Observational entropy: entropy in the context of indistinguishability operators

Enric Hernàndez
Secció de Matemàtiques i Informàtica
ETSAB. Avda. Diagonal 649, 08028 Barcelona
Universitat Politècnica de Catalunya
enriche@lsi.upc.es

Jordi Recasens
Secció de Matemàtiques i Informàtica
ETSAV. Pere Serra 1-15, 08190 Sant Cugat
Universitat Politècnica de Catalunya
recasens@ea.upc.es

Summary

This work is intended to study the notion of entropy in the context of indistinguishability operators. The adapted notion of entropy should combine two kinds of uncertainty: the one coming from random (probabilistic uncertainty) and the uncertainty in the observation of the elements.

Keywords: entropy, observational entropy, Shannon's entropy, observation degrees, T-indistinguishability operators.

1 Introduction

The problem of measuring the uncertainty of a set of events is not new. The first attempts tried to quantify the uncertainty associated to a random experiment. So, Hartley (1928) captured the intuitive idea that the more possible results for an experiment, the less it can be predicted, defining the uncertainty as $\log(n)$, where n is the number of possible results. Then we have $\log(n) < \log(n+1)$, and also that the uncertainty associated with the product of two experiments is the sum of the single uncertainties ($\log(pq) = \log(p) + \log(q)$).Anyway, this measure had the drawback of ignoring the probability of the events. This difficulty was overcome by Shannon (1948) defining the entropy of a random variable as:

$$H(X) = -\sum_{x \in X} p(x) \log_2 p(x) \qquad (1)$$

It is important to note that this measure was thought within the frame of communication theory, specifically for facing issues concerning channel reliability and reduction of transmission cost, but ignoring the semantic content of the messages involved. What happens when events "carry" a concrete meaning, defined in terms of risk, utility, ..? [2] presents a situation worthwhile enough to be reproduced, in order to make clear that point: let us suppose two roads R_1,R_2 with probabilities of accident 0.28 and 0.72 respectively. Shannon's entropy is the same for both of them, but it seems quite obvious which one we would choose, based on the uncertainty of the risk of an accident. Therefore, providing the set of events with a particular semantics requires a "further step", in the sense that we need to adapt Shannon's measure in order to express random uncertainty in terms of this semantics. In this work, this semantics will be established by defining an indistinguishability relation among the elements of some domain, making some elements indistinguishable from others. The main idea is that the occurrence of two different events, but indistinguishable by the indistinguishability relation defined, will count as the occurrence of the same event when measuring the "observational" entropy.

Let us recall the definition of T-indistinguishability operator [3]:

Definition 1.1: A fuzzy relation E defined over X is a T-indistinguishability operator (being T a t-norm) if $\forall x, y, z \in X$:

- $E(x,x) = 1$ (reflexivity)
- $E(x,y) = E(y,x)$ (symmetry)
- $T(E(x,y), E(y,z)) \leq E(x,z)$ (T-transitivity)

2 Previous work

There has been done some work intended to combine random and fuzzy uncertainty. The proposals can be grouped in three main tendencies. In the sequel $P(X)$ will denote a probability distribution on X and A a fuzzy set on X with membership function μ_A.

- "Global" combination: algebraic combination (sum, product, ..) of the two kinds of uncertainty which, in turn, are calculated isolatedly. An exemple is the measure of Xie and Bedrosian (1984) [4]:

$$H(A,P) = -\sum_{x\in X} p(x)\log_2 p(x) - \sum_{x\in X}(\mu_A(x)\log_2\mu_A(x) + (1-\mu_A(x))\log_2(1-\mu_A(x))) \quad (2)$$

, where the first term measures the random entropy (Shannon's entropy), and the rest measures the fuzziness degree of A (by using DeLuca and Termini's (1972) expression).

- "Local" combination (element by element): each element combines its contribution to both kinds of uncertainty locally. The final value is the sum of all those individual contributions. An exemple is the DeLuca and Termini's measure:

$$H(A,P) = -\sum_{x\in X} p(x)\,[\mu_A(x)\log_2\mu_A(x) + (1-\mu_A(x))\log_2(1-\mu_A(x))] \quad (3)$$

- Weighted measures: one of the two values (probability or possibility degree) "weights" the contribution of the other. Zadeh's measure [6] exemplifies this trend:

$$H(A,P) = -\sum_{x\in X}\mu_A(x)p(x)\log_2 p(x) \quad (4)$$

,or the one found in [2]:

$$H(A,P) = -\sum_{x\in X}\frac{\mu_A(x)}{\sum_{x\in X}p(x)\mu_A(x)}p(x)\log_2 p(x) \quad (5)$$

3 Observational entropy

In [5] a new measure of entropy, suitable to operate on domains over which a similarity relation has been defined, is introduced. Our point of view is that this measure overcomes some limitations of the previous proposals. In this section we will recall some basic definitions and properties of this entropy measure, and will provide a new interpretation in terms of the observability of the elements.

Definition 3.1: Let E a T-indistinguishability operator on a set X. The observation degree of $x_j \in X$ is defined by:

$$\pi(x_j) = \sum_{x\in X} p(x)E(x,x_j). \quad (6)$$

Due to the reflexivity of E, this expression can be rewritten as:

$$\pi(x_j) = p(x_j) + \sum_{x\in X|x\neq x_j} p(x)E(x,x_j). \quad (7)$$

This definition has a clear interpretation: the possibility of observing x_j is given by the probability that x_j really happens (expressed by the first term), plus the probability of ocurrence of some element "very close" to x_j, weighted by the similarity degree . In other words, the first term measures the possibility of really observing x_j, while the second term measures the possibility of observing x_j mistakenly (x_j didn't really happen). It holds:

Lemma 3.2:

$$\forall x \in X : 0 \leq \pi(x) \leq 1. \quad (8)$$

Corollary 3.3:

$$0 \leq \sum_{x\in X}\pi(x) \leq Card(X). \quad (9)$$

It should be noted that $\pi(X)$ is not a probability distribution since $\sum_{x\in X}\pi(x) \neq 1$.

Definition 3.4: The quantity of information received by observing x_j is defined by:

$$C(x_j) = -\log_2\pi(x_j). \quad (10)$$

Definition 3.5: Given a T-indistinguishability operator E on X, and P a probability distribution on X, the observational entropy (HO) of the pair (E,P) is defined by:

$$HO(E,P) = \sum_{x\in X} p(x)C(x). \quad (11)$$

Let's suppose the following case: $X = \{x_1, x_2\}$, P the following probability distribution $p(x_1) = p(x_2) = 0.5$, and E the T-indistinguishability operator defined by $E(x_1,x_1) = E(x_2,x_2) = 1$; $E(x_1,x_2) = E(x_2,x_1) =$

0. It is trivial to check that $HO(E,P) = 1$. This result suggests the following definition:

Definition 3.6: The information received by observing an event between two equally probable and fully distinguishable events will define the unit of measure for the observational entropy: the observable bit.

Lemma 3.7: Let E be a T-indistinguishability operator on X, P a probability distribution on X and $H(P)$ the Shannon's entropy.

$$HO(E,P) \leq H(P). \tag{12}$$

Lemma 3.8: Let E, E' be two T-indistinguishability operators on X and P a probability distribution on X.

$$E \leq E' \rightarrow HO(E,P) \leq HO(E',P). \tag{13}$$

Lemma 3.9: Given $X = \{x_1, x_2\}$, E the T-indistinguishability operator on X such that

$$E(x_i, x_j) = \begin{cases} 1 & , i = j \\ 0 & , i \neq j \end{cases}$$

and P a probability distribution on X, then:

$$HO(E,P) = H(P). \tag{14}$$

4 Observation degree as expected value of a random variable

In this section, a new interpretation of the observation degree is given. This degree was defined as:

$$\pi(x_j) = \sum_{x \in X} p(x) E(x, x_j).$$

When working on discrete domains, the T- indistinguishability operator can be represented by a symmetric matrix M, where the (i,j) component takes the value $E(x_i, x_j)$. A closer look to the columns (or rows) of matrix M, leads us to realize that column i contains the indistinguishability degrees of all the elements with respect to the element x_i. Therefore, we can define the fuzzy set "similarity with x_i" ($\approx_i$) on X, also called singletons or simply columns on the literature, as:

$$\forall x_j \in X \quad : \quad \approx_{x_i}(x_j) = E(x_i, x_j). \tag{15}$$

Now, we define the random variable $G_{\approx_{x_i}}$ over the interval [0,1] with the following probability distribution $P_{\approx_{x_i}}$:

$$\forall r \in [0,1] \quad : \quad P_{\approx_{x_i}}(r) = \sum_{x \in X | \approx_{x_i}(x) = r} p(x). \tag{16}$$

It is easy to see that $P_{\approx_{x_i}}$ is a probability distribution over the set of membership degrees ([0,1]), where each $r \in [0,1]$ takes as a probability, the sum of all the elements whose "similarity degree" with x_i is just r. Finally, the next equality holds:

$$\forall x \in X \quad : \quad \pi(x) = E(G_{\approx_x}). \tag{17}$$

, that is, the observation degree of an element x is the expected value of the random variable $G_{\approx_x}$. In other words, the observation degree is the expected value for the "similarity degrees with x".

5 Simultaneous observation degree

Arrived at this point, we will introduce the concept of simultaneous observation degree. The meaning of this new degree could be stated as follows: given that an indistinguishability relation has been defined, it is possible for two independent observers to disagree in the observation of an event. For instance, observer A may have observed event x_i, while observer B may have observed event x_j, if x_i and x_j are "close enough". Obviously, if the events were fully distinguishable, this "overlapping" or "simultaneous observation" could not have been possible (assuming the abssence of noise or error).

Definition 5.1: The simultaneous observation degree of the subset $\{x_1, \ldots x_k\}$ is defined by:

$$\begin{aligned} O(\{x_1, \ldots, x_k\}) &= E(G_{[\approx_{x_1} \wedge \ldots \wedge \approx_{x_k}]}) \\ &= E(G_{T(\approx_{x_1}, \ldots, \approx_{x_k})}). \end{aligned} \tag{18}$$

That is, the simultaneous observation degree of $\{x_1, \ldots, x_k\}$ is the expected value of the similarity degrees with respect to $x_1, \ldots, x_k$.

Lemma 5.2:

$$\begin{aligned} O(\{x_1, \ldots, x_k\}) &= 1 \leftrightarrow \\ \forall x_i, x_j \in X \quad &: \quad E(x_i, x_j) = 1 \end{aligned} \tag{19}$$

Lemma 5.3:

$$O(\{x_1,\ldots,x_k\}) = 0 \leftrightarrow \forall x_i \in X \mid p(x_i) > 0 : \exists x_j \mid E(x_i,x_j) = 0 \quad (20)$$

6 Conditional observation degree

In the last section we dealt with the scene in which there was disagreement among observers "equipped" with the same indistinguishability relation. Now, we will examine the case of observers with different indistinguishability relations. Here, the main question is: How does affect the observations performed by an observer in order to predict the observations of some other observer?. For instance, let us suppose that we know that observer A using indistinguishability E_A has observed event x_i. This fact restricts the events that really might have been happened to the set of events "close enoug" to x_i with respect to E_A . This restriction in the set of possible events affects to the observability of observer B.

Definition 6.1: Let E, E' two T-indistinguishability operators on a set X, P a probability distribution on X and T a t-norm. The conditional observation degree of event x_j in (E,P), given that x_i has been observed in (E',P) is defined by:

$$\pi_{E|E'}(x_j \mid x_i) = \sum_{x\in X} p(x)T(E'(x_i,x),E(x,x_j)). \quad (21)$$

Lemma 6.2:

$$\pi_{E|E}(x_j \mid x_i) = O_E(\{x_i,x_j\}). \quad (22)$$

That is, when the T-indistinguishability operators are equal, the conditioned observation degree of x_j, given the observation x_i, is the simultaneous observation degree between x_i and x_j in E. It should be noted that the conditioned observation degree is not necessary symmetric, but when $E = E'$, symmetry holds.

7 Conclusions and future work

In this work we have studied the problem of adapting the notion of entropy when, besides the probabilistic uncertainty, we have the uncertainty coming from a indistinguishability relation defined over the set of events. New interpretations are given, which allow us to define the "observe" paradigm. This paradigm leads to the definition of some "observation" degrees, and to study their relationship with the concept of entropy. As future work we propose the next points:

- Joint observational entropy and conditional observational entropy.
- Extend the results to continuous domains .
- Application to the construction of fuzzy decision trees and clustering.

Acknowledgements

This work has been funded by the project PB94-1208 of the DGICYT.

References

[1] Fazlollah Reza. An introduction to information theory. *Dover Publications, inc. New York.*

[2] Pedro Gil (1981) Álvarez. Teoría matemática de la información. *Ediciones ICE.*

[3] J. Jacas and J. Recasens (1995). Fuzzy T-transitive relations: eigenvectors and generators. *Fuzzy sets and systems 72.* pp 147-154.

[4] Nikhil R. Pal and James C. Bezdek (1994). Measuring fuzzy uncertainty. *IEEE Transactions on Fuzzy Systems, Vol 2, No 2* pp 107-118.

[5] Ronald R. Yager (1992). Entropy measures under similarity relations. *Int. J. General Systems, Vol 20.* pp 341-358.

[6] L. Zadeh (1968). Probability measures of fuzzy events. *J. Math. Analysis and Applicat. Vol 23.* pp 421-427.

Inmersión de W-particiones en un espacio producto

Jordi Recasens
Secció Matemàtiques i Informática
E.T.S.A.V.
Universidad Politécnica de Cataluña
jordi.recasens@ea.upc.es

Adolfo R. De Soto
Area de Lenguajes y Sistemas Informáticos
Facultad de Económicas y Empresariales
Universidad de León
ddears@unileon.es

Resumen

Se muestra un método para realizar una inmersión de una W-partición cualquiera en una clase de particiones producto que son generadas por familias de particiones en cada factor. Estas familias admiten una interpretación sencilla como variables lingüísticas.

Palabras clave: Particiones borrosas, indistinguibilidades, variables lingüísticas, sistemas de reglas borrosos.

1 Introducción

El desarrollo actual de los sistemas de información provoca que toda organización disponga de una cantidad ingente de datos, tanto estructurados como no estructurados, y que considere esta información como un elemento estratégico. Pero un almacén de datos no es en sí mismo un almacén de conocimiento; el conocimiento es fundamentalmente sintético. Así a medida que se recogen más datos, más necesario se hace el disponer de métodos automáticos de extracción de información relevante y resumida para poder obtener algún provecho real de los mismos.

Los sistemas basados en reglas borrosas tienen varias propiedades que les hacen especialmente adecuados para esta tarea. En primer lugar son aproximadores universales por lo que pueden realizar aproximaciones a cualquier función con una precisión arbitraría. En segundo lugar son sistemas con una interpretación sencilla y fácilmente implementables. Esto se corroborá en la facilidad con la que un experto puede declarar su conocimiento mediante un sistema de este tipo.

Extraer conocimiento de una fuente de datos es el proceso inverso al que se realiza cuando se diseña un sistema de reglas borrosas con la ayuda de un experto: de un conjunto de datos debemos extraer información para el experto. La realidad nos muestra que los métodos automáticos de extracción de reglas llevan a sistemas donde los conjuntos borrosos obtenidos son "extraños", siendo en muchas ocasiones simples funciones de un universo en $[0,1]$ más que funciones de pertenencia de algún predicado borroso conocido o interpretable. Esto no supone ningún problema si lo que se desea es extraer reglas para ser usadas posteriormente en sistemas de control automático, la información la obtiene la máquina y la usa la máquina; pero cuando estos métodos están orientados a ofrecer información al experto humano la importancia de poder realizar un interpretación sencilla de las reglas es vital.

El gran logro de los sistemas basados en reglas borrosos es contar con un medio que traduce información en lenguaje humano a información en el lenguaje de una máquina. Si se desean conseguir métodos que infieran conocimiento útil para el hombre a partir de datos, debe ser la máquina, es decir los mecanismos utilizados, la que se acerque al lenguaje del hombre. Por eso consideramos que estos métodos deben tener como principal objetivo obtener información interpretable. La pregunta es evidentemente si esto se puede conseguir manteniendo el carácter de aproximadores universales de estos sistemas.

En el trabajo [1] se introdujo una familia de W-particiones sobre $[0,1]$, $P = \{P_n\}_{n\in\mathbb{N}}$, donde cada partición, $P_n = \{\mu_{r:n}\}_{r\in 0\ldots n}$, está formada por los conjuntos borrosos $\mu_{r:n}(x) = \max(0, 1 - n\left|x - \frac{r}{n}\right|)$. Cada partición P_n es la única partición compatible, en el sentido dado posteriormente, con la potencia n-ésima de la indistinguibilidad de Lukasiewicz definida como $E_W^n(x,y) = \max(0, 1 - n\,|x-y|)$. La importancia de esta asociación es que la indistinguibilidad es una medida de la granularidad, entendida como capacidad de distinción, de la partición P_n : a medida que n crece cada partición P_n está formada por más conjuntos borrosos y la indistinguibilidad da valores menores para cada par de puntos x, y. Además la familia P admite una interpretación como variable lingüística, presen-

tando una relación de antonimia entre los conjuntos borrosos de cada partición P_n y disponiendo de un conjunto de modificadores lingüísticos que permiten pasar de una partición a otra. El uso de modificadores en el lenguaje permite precisamente una mayor capacidad de distinción mediante el aumento del número de etiquetas lingüísticas.

En este trabajo se continua investigando las propiedades de las W-particiones iniciado en [1] y de las W-particiones P_n como ejemplos especialmente interesantes. En concreto se prueba que a toda W-partición se le puede asociar una W-indistinguibilidad de tal forma que la partición es compatible con ella. De este modo se relacionan dos mundos, el de relaciones de indistinguibilidad y de particiones hasta ahora separados. Además se estudia el paso de las particiones P_n al universo producto.

2 Definiciones previas

Se define una T-partición, con T una t-norma, como una familia de conjuntos borrosos $H = \{\mu_1, \ldots, \mu_n\}$ sobre un universo X que verifica las dos condiciones siguientes:

1. $T(\mu_i(x), \mu_j(x)) = 0$ para todo par $i, j \in 1, \ldots, n$ con $i \neq j$;
2. $T^*(\mu_1(x), \ldots, \mu_n(x)) = 1$ para todo $x \in X$, siendo T^* la t-conorma dual de T respecto a una negación fuerte.

Las hard-particiones definidas por Bezdek [4] y que se obtienen al utilizar el algoritmo de las c-medias son un caso particular de las W-particiones, con $W(x, y) = \max(0, x + y - 1)$ la t-norma de Lukasiewicz.

Dada una T-indistinguibilidad $E : X \times X \to [0, 1]$, se dice que una T-partición H es compatible con ella si

1. $T(\mu_i(x), \mu_i(y)) \leq E(x, y)$,
2. existe un $a_i \in X$ con $E(x, a_i) \leq \mu_i(x)$,

para todo $x, y \in X$ e $i \in 1 \ldots n$.

En [1] se probó que los subconjuntos borrosos $\mu_{r:n}(x) = \max(0, 1 - n\left|x - \frac{r}{n}\right|)$ para $r = 0, \ldots, n$ forman una W-partición sobre $[0, 1]$ que denotaremos por $P_n = \{\mu_{r:n}\}_{r=0,\ldots,n}$. De hecho P_n es la única W-partición compatible con la indistinguibilidad E_W^n obtenida al operar n veces la indisitnguibilidad de Lukasiewicz consigo misma.

El siguiente resultado muestra que puede realizarse la operación inversa.

Teorema 1 *Dada una hard-partición $H = \{\mu_1, \ldots, \mu_m\}$ sobre un universo X es posible construir una W-indistinguibilidad compatible con H.*

Demostración. Consideramos el universo $\overline{X} = X \cup \{a_1, \ldots, a_n\}$ y extendemos los conjuntos borrosos de la partición mediante la expresión $\mu_i(a_j) = \delta_{ij}$. Cada μ_i define una indistinguibilidad E_{μ_j} en $\overline{X}$ y el mínimo de todas $E = \min(E_{\mu_1}, \ldots, E_{\mu_m})$ es una W-indistinguibilidad compatible con H. Para ver esto último sólo hay que comprobar que los μ_i son columnas de E y lo son ya que

$$\begin{aligned} E(a_i, x) &= \min_{j=1,\ldots,m} E_{\mu_j}(a_i, x) \\ &= \min_{j=1,\ldots,m} 1 - |\mu_j(a_i) - \mu_j(x)| \\ &= \mu_i(x) \\ &= E_{\mu_i}(a_i, x). \end{aligned}$$

De todas las igualdades la única que no es por definición es la tercera y se prueba teniendo en cuenta que si $j = i$ entonces $E_{\mu_j}(a_i, x) = 1 - |\mu_i(a_i) - \mu_i(x)| = \mu_i(x)$ y si $j \neq i$ entonces $E_{\mu j}(a_i, x) = 1 - |\mu_j(a_i) - \mu_j(x)| = 1 - \mu_j(x)$ y dado que $\sum_{k=1}^{m} \mu_k(x) = 1$, se tiene que $\mu_i(x) \leq 1 - \mu_j(x)$. ■

El teorema anterior requiere la introducción artificial de prototipos para cada clase de la partición. Si las clases son conjuntos borrosos normalizados esta operación no es necesaria.

3 Particiones producto cartesiano

Dadas dos T-indistinguibilidades E, F sobre universos X, Y respectivamente, se define la T-indistinguibilidad producto cartesiano como

$$(E \times F)(\mathbf{z}_1, \mathbf{z}_2) = \min(E(x_1, x_2), E(y_1, y_2))$$

siendo $\mathbf{z}_1 = (x_1, y_1)$, $\mathbf{z}_2 = (x_2, y_2)$ puntos pertenecientes al universo producto $X \times Y$. La definición puede ser ampliada a cualquier dimensión finita.

De una forma similar, podemos construir una T-partición producto a partir de un número finito de T-particiones. Para el caso de dos, sean $H_X = \{\mu_1, \ldots, \mu_n\}$, $H_Y = \{\nu_1, \ldots, \nu_m\}$ dos T-particiones de los universos X eY respectivamente. Se define la partición producto $H_X \times H_Y$ como la clase formada por los conjuntos borrosos:

$$\rho_{ij}(\mathbf{z}) = \min(\mu_i(x), \nu_j(y))$$

con $\mathbf{z} = (x, y) \in X \times Y$.

Teorema 2 *La partición producto $H_X \times H_Y$ es una W-partición compatible, con la W-indistinguibilidad*

producto, sobre $X \times Y$ si H_X y H_y lo son sobre X e Y respectivamente.

Considerando el producto cartesiano $[0,1]^n$ y en cada factor una indistinguibilidad del tipo $E_W^n(x,y) = \max(0, 1 - n|x-y|)$ con $n \in \mathbb{N}$., la indistinguibilidad producto cartesiano tiene asociada una partición producto compatible. Esta partición proviene en cada factor de particiones que tienen una fácil interpretación como variables lingüísticas y que además, en caso necesario, pueden ser refinadas utilizando modificadores lingüísticos. Por tanto, en cada factor tenemos una variable lingüística generada por dos predicados contrarios, con lo que el esquema de conocimiento que inducen sobre la variable es similar al que utiliza el hombre para clasificar muchas variables.

4 Inmersión de W-particiones

Definición 3 *Dados un conjunto X con una particiones $H_1 = \{\mu_1, \ldots, \mu_n\}$ y otro conjunto Y con partición $H_2 = \{v_1, \ldots, v_m\}$, se dice que ambas particiones son homométricas si existe una aplicación $f : X \to Y$ tal que a todo $\mu_i \in H_1$ se le asocia un conjunto borroso $f_*(\mu_i) \in H_2$ tal que*

$$\mu_i(x) = (f_*(\mu_i))(f(x)).$$

El siguiente teorema prueba que toda W-partición sobre un universo finito es homométrica a una partición producto generada por particiones P_n en cada factor, de tal forma que se obtiene un marco unificado para su estudio.

Teorema 4 *Sea X un conjunto y H una W-partición de X en c clases. X es homométrico a algún $[0,1]^n$ con la partición generada por el producto cartesiano de las indistinguibilidades $E_W^{p_i}$ en cada factor. Se puede tomar $n \leq c$*

Demostración. Sea $H = \{\mu_1, \ldots, \mu_c\}$ la partición y consideremos la aplicación

$$f : X \to [0,1] \times \ldots \times [0,1]$$

definida como

$$f(x) = (\mu_1(x), \ldots, \mu_c(x)).$$

Sobre cada factor $[0,1]$ sea E_W la indistinguibilidad de Lukasiewicz y $E_W \times \overset{c}{\ldots} \times E_W$ la indistinguibilidad producto. Definiendo

$$f_*(\mu_i) = \min(1 - Id, \ldots, Id, \ldots, 1 - Id)$$

siendo Id la aplicación identidad, tenemos una homometría puesto que

$$\begin{aligned} f_*(\mu_i)(f(x)) &= \min(1 - \mu_1(x), \ldots, \mu_i(x), \ldots, 1 - \mu_c(x)) \\ &= \mu_i(x) \end{aligned}$$

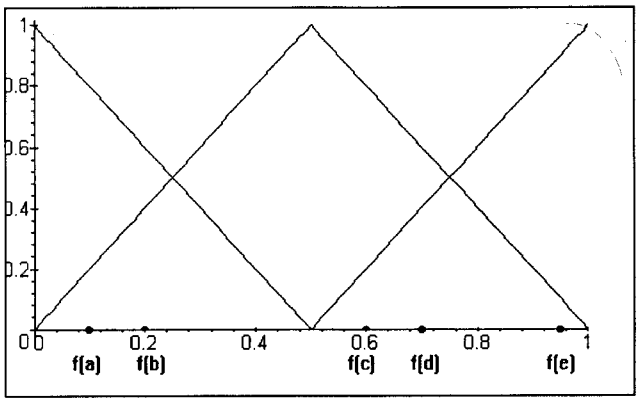


Figura 1: Inmersión por una homometría

ya que para todo $j \neq i$ se tiene que $\mu_i(x) \leq 1 - \mu_j(x)$. ■

El siguiente ejemplo muestra que puede que el número de factores de la inmersión sea menor que el número de clases de la partición.

Ejemplo 5 *Consideremos $X = \{a,b,c,d,e\}$ y $H = \{\mu_1, \mu_2, \mu_3\}$ con*

$$\begin{aligned} \mu_1 &= 0,8/a + 0,6/b + 0/c + 0/d + 0/e \\ \mu_2 &= 0,2/a + 0,4/b + 0,8/c + 0,4/d + 0,1/e \\ \mu_3 &= 0/a + 0/b + 0,2/c + 0,6/d + 0,9/e \end{aligned}$$

Tomemos la partición asociada a E_W^2 que estará formada por tres clases $P_2 = \{\mu_{0:2}, \mu_{1:2}, \mu_{2:2}\}$. Definiendo la función f como

$$\begin{aligned} f(a) &= 0,1 \quad f(b) = 0,2 \quad f(c) = 0,6 \\ f(d) &= 0,7 \quad f(e) = 0,95 \end{aligned}$$

y $f_*(\mu_i) = \mu_{i-1:3}$ tenemos una homometría, con lo que hemos conseguido una inmersión en un producto cartesiano de dimensión $n = 1$. (Veáse la figura 1)

Este ejemplo sugiere el concepto de dimensión de una W-partición como el mínimo número de factores para el que existe una homometría.

Definición 6 *Con las notaciones del Teorema 4, diremos que (X,H), o más brevemente H, tiene dimensión n si, y sólo si, X es homométrico a algún $[0,1]^n$ pero no a $[0,1]^{n-1}$.*

Esta dimensión está acotada por el número de clases de la partición y puede considerarse como una medida del solapamiento que presentan las clases sobre el universo X. Parece difícil hallar algoritmos eficaces para calcular la dimensión de una W-partición dada. Las siguientes ideas proporcionan cotas razonables. Por ejemplo, para que sea unidimensional, es

necesario que cada elemento pertenezca a lo sumo a dos clases, pero esto no es suficiente, porque podría haber particiones " "cerradas" o "cíclicas": x_1 es de las clases c_1 y c_2, x_2 de las c_2 y c_3, y x_3 de las c_3 y c_1, por ejemplo. De hecho, si consideramos un grafo en el que los vértices se correspondan con las clases y dos vértices estén unidos si y sólo si existe algún elemento que pertenezca a ambas clases, entonces la partición es unidimensional si y sólo si el grafo es un subgrafo de una cadena no cerrada. Más general, para que una partición sea n-dimensional, es necesario que cada elemento pertenezca a lo sumo a $2n$ clases, lo que se traduce en que de cada vértice salgan a lo sumo $2n-1$ aristas. Una vez más, esto no es suficiente, porque las clases que quedaran a los lados no pueden conectarse entre sí.

Definición 7 *Diremos que un grafo es una malla si, y sólo si, es isomorfo al siguiente grafo. Tomamos como vértices los puntos* $(x_1, x_2, \ldots, x_n)$ *de* $\mathbb{R}^n$ *con coordenadas enteras y tales que para todo* i *existe* c_i *con* $0 \leq x_i \leq c_i$. *Dos vértices estarán unidos por una arista si, y sólo si, cada coordenada se diferencia en a lo sumo en una unidad.*

En el caso bidimensional, los vértices de una malla formarán cuadrados y las aristas serán sus lados y diagonales. Con esta definición se tienen los siguientes resultados.

Teorema 8 *Una W-partition es n-dimensional si, y sólo si, es isomorfa a un subgrafo de una malla de* R^n, *pero no de* R^{n-1}.

Corolario 9 *Si cada elemento pertenece a lo sumo a* $2n$ *clases, la W-partition es a lo sumo n-dimensional.*

Demostración. $2n-1$ es el número de aristas que salen de los vértices situados en las esquinas de una malla.

5 Conclusiones

Estos resultados pueden verse como los primeros estudios destinados a conocer en que circunstancias una partición puede aproximarse como un producto cartesiano de particiones del tipo P_n. El lograr esta aproximación permitiría un medio para lograr una interpretación sencilla de las condiciones que deben cumplirse en cada factor para que un elemento forme parte de una clase de la partición. Estas condiciones pueden escribirse en forma de reglas por lo que se podrían tener métodos automáticos de extracción de reglas usando conjuntos borrosos que admiten una interpretación como etiquetas lingüísticas de una variable lingüística. Además puede demostrarse fácilmente, basándose en el trabajo[3], que el uso de las particiones P_n en cada variable de entrada conducen a sistemas de reglas que son aproximadores universales.

Referencias

[1] De Soto, A.R.; Recasens, J. "Particiones, Indistinguibilidades y variables lingüísticas". VIII Congreso Español sobre Tecnologías y Lógica Fuzzy, pp. 245–250, Pamplona, 1998.

[2] De Soto, A.R.; Trillas, E. "Second Thoughts on Linguistic Variables". Proc. of 18th Inter. Conference of the North American Fuzzy Information Processing Society – NAFIPS'99, pp. 37–41, New York.

[3] Castro, J.L. "Fuzzy Logic Controllers are Universal Approximators" IEEE Trans. on Syst., Man and Cyber., 25–4, pp. 629–635, 1995.

[4] Klir, G.J.; Yuan, B. "Fuzzy Sets and Fuzzy Logic. Theory and Applications" Prentice Hall, New Yersey, 1995.

[5] Zadeh, L.A. (1976) "The Concept of a Linguistic Variable and its Application to Aproximate Reasoning. I, II and III" en *Fuzzy Sets and Aplications. Selected Papers by L.A. Zadeh* editado por Yager, R.R. et al. pp. 219–366, John Wiley and Sons, New York.

ON THE CONSTRUCTION OF M-TRANSITIVE RELATIONS.

Joan Jacas
Sec. de Mat.bi Inf.ETSAB.
Universitat Politècnica de Catalunya
Av.Diagonal 649, 08028 Barcelona.Spain.
Jacas@ea.upc.es.

Jodi Recasens.
Sec. Mat i Inf. ETSAV.
Universitat Politècnica de Catalunya.
Pere Serra 1-15, 08190 Sant Cugat. Spain.
Recasens@ea.upc.es

Abstract

The transitivity of fuzzy relations with respect to a subset of quasi-linear functions in the unit interval is studied. A new construction of transitive relations with respect to an archimedean t-norm is proposed.
Keywords: fuzzy relation, transitivity, quasi-inverse, fuzzy connective

1. INTRODUCTION[1]

All the branches of human knowledge have the need to distinguish the objects they are working with, therefore it is necessary to define some type of "equality". From a classical point of view, this equality is an equivalence relation defined on the objects of the universe of discourse. Nevertheless, an increasing number of problems arising mainly from applied sciences and specially from the so-called soft-sciences, point out at the introduction of equalities that are not transitive in the usual sense [11,10,12]. A new approach is needed. This new point of view is obtained by introducing the concept of "degree of equality or similarity". This setting leads to a definition of equality linked with the idea of "proximity" having a clear topological flavour. The first step is to generalise the concept of binary relation by introducing the fuzzy relations in the universe of discourse X. In this context the definition of the transitive property plays a key role [7, 6,14, 15]. A fuzzy relation R defined on a set X fulfils some kind of transitivity if, given three elements x, y, z of X it is possible to calculate a lower bound of the value $R(x, z)$ given the values of $R(x, y)$ and $R(y, z)$. In a general framework this calculus is carried out by means of binary operation $M:[0,1]\times[0,1] \rightarrow [0,1]$ that can be interpreted as a conjunction [3] in the multivalued logic underlying the fuzzy sets theory. Therefore it seems natural to suppose that is a commutative operation. Further, if we want to include in this schema the classical transitive relations, then $M(1,1)=1$. Summing up, a fuzzy relation R defined in X is M-transitive if there is a commutative operation $M:[0,1]\times[0,1] \rightarrow [0,1]$ such that $M(1,1)=1$ and

$$M(R(x, y), R(y, z)) \leq R(x, z)$$

Under these hypotheses, several authors [3,15 ,7,8,5] have published a large number of papers devoted to this type of relations taking the operation M within the set T of t-norms. In this paper we study some aspects of the relations that are transitive with respect to operations that are more general relations than the t-norms such as the quasi-linear functions [1].

2. PRELIMINARIES

Throughout this paper I will represent the unit interval [0, 1] and M a mapping $I\times I \rightarrow I$ fulfilling the following conditions:

2.1 $M(1, 1)=1$.
2.2 $M(x, y) = M(y, x)$ for all $x, y \in I$
2.3 M is non-decreasing with respect to each variable.
2.4 M is continuous.

Definition 2.1
A fuzzy relation R defined on a set X is M-transitive if

$$M(R(x, y), R(y, z)) \leq R(x, z),$$

for all x, y, z of X.

Definition 2.2
Given $(x, y)\in I \times I$, the *residuation* of (x, y) with respect to M is the number

$$\hat{M}(x \mid y) = Sup \{\alpha \in [0,1] \mid M(\alpha, x) \leq y\} \qquad (1),$$

if it exists.
Given $A \subset I\times I$, if for all $(x, y)\in A$ there exists $\hat{M}(x \mid y)$ then, we can define a new operation $\hat{M} : A \rightarrow [0,1]$ given by the formula (1).
This operation will be termed as the *quasi-inverse* of M in A.
All the preceding definitions are very general, from now on we will focus our attention to a particular set of M-operations, the subset of the quasi-linear functions [1] defined by

$$M(x, y) = f^{[-1]}(p f(x) + q f(y)) \qquad (2).$$

Where $p, q \in I$ and $f: I \rightarrow [0,+\infty]$ is a strictly monotonous function [1,2].

[1] Partially supported by the CICYT project TIC96-1393-C06-05

The conditions 2.1-4 fulfilled by the operation M impose the following restrictions to p, q and f:

2.7 $M(1,1) = 1 \Rightarrow f(1) = 0$ or $f(1) = +\infty$.

2.8 M commutative $\Rightarrow p = q$.

2.9 If M is interpreted as a conjunction [4], then $M(x,y) \leq (x+y)/2$ for all $(x, y) \in I \times I \Rightarrow f$ is a strictly decreasing function with $f(1)=0$ and $p \geq 1/2$.

In the formula (2), $f^{[-1]}$ represents the pseudo-inverse of f in the usual sense. Nevertheless, in this paper it will be used in more general way.

Definition 2.3

Given an strictly decreasing function $f:[0,1] \rightarrow [0,+\infty]$, the pseudo-inverse of f is the function

$$f^{[-1]}:[-\infty,+\infty] \rightarrow [0,1]$$

defined in de following way:

$$f^{[-1]}(a) = \begin{cases} Inf\, f^{-1}([0,a[) & a \geq 0 \\ 1 & a < 0 \end{cases}$$

The pseudo-inverse defined in this way is always a continuous function and coincides with its unique quasi-verse [8], therefore $f^{[-1]} \circ f = id_{I0,1I}$ and $\left[f^{[-1]}\right]^{[-1]} = f$ if f is continuous.

Summing up, $M(x,y) = f^{[-1]}(p(f(x)+f(y)))$, where f is a strictly decreasing continuous function, $f^{[-1]}$ its pseudo-inverse and $1 \geq p \geq 1/2$

The set of this type of functions will be denote by $\boldsymbol{M}$. Each function is determined by a unique f, except for constant factors, and a value $p \in I$.

The operation M is called *strict* or *non-strict* depending on whether $f(0) = +\infty$ or $f(0) = k < +\infty$ respectively.

The set $\boldsymbol{M}$, even being restricted to aforementioned conditions, contains a vast number of operations including the Archimedean t-norms (p=1) and the quasi-arithmetic means (p=1/2). In particular, for $f(x)=-lnx$, $p=1/2$ we have $M(x,y) = \sqrt{xy}$ (geometric mean) and for $f(x)= 1-x$, $p=1/2$ we have $M(x,y) = \dfrac{x+y}{2}$ (arithmetic mean)

If $M \in \boldsymbol{M}$ is generated by f, the t-norm $T_f(x,y) = f^{[-1]}(f(x)+f(y))$ will be termed the t-norm associated to the family

$$M_f = \{M \mid M(x,y) = f^{[-1]}(p(f(x)+f(y)); p \in [1/2,1]\}$$

3. TRANSITIVITY AND RESIDUATION.

In this section we study the relation between M–transitive relations ($M \in \boldsymbol{M_f}$) and the fuzzy relations generated by using the residuation of M.

Proposition 3.1

Given $M \in \boldsymbol{M_f}$ *if for* $(x,y) \in I \times I$ there exists $\hat{M}(x \mid y)$ then

$$\hat{M}(x \mid y) = f^{[-1]}\left(\frac{f(y)}{p} - f(x)\right)$$

In order to study which operations of the set $\boldsymbol{M}$ admit a residuation we need the following lemmas.

Lemma 3.1

If $M_f \in \boldsymbol{M}$ is strict then for all $x \in [0,1]$,

$$[0,x] \subset Im\, M_f(\cdot, x)$$

Lemma 3.2

Let $M \in \boldsymbol{M_f}$ be a non-strict function with $f(0)=k$. If $p \geq 1/2$ then, for all $x \in \left[0, f^{-1}\left(\left(\frac{1-p}{p}\right)k\right)\right]$ we have $[0,x] \subset Im\, M(.,x)$

From the preceding lemmas it immediately follows.

Theorem 3.1

If $M \in \boldsymbol{M_f}$ is strict, then M has a quasi-inverse $\hat{M}$ in the unit interval.

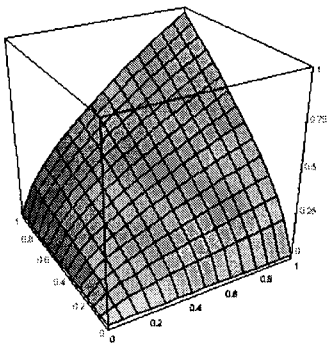


Figure 1

Example 3.1

If $f(x)=-\ln(x)$ then $f^{[-1]}(x) = Max(e^{-x},1)$, $M(x,y) = \sqrt[p]{x \cdot y}$ (figure 1) and

$$\hat{M}(x \mid y) = Min\left(\frac{y^{\frac{1}{p}}}{x},1\right) \quad (p = 2/3).$$ Figure 2.

Theorem 3.2

Given a non-strict $M_f \in \boldsymbol{M}$ with $f(0)=k$, $p>1/2$ and $\lambda = f^{-1}\left(\frac{1-p}{p}\right)k$, M_f has a quasi-inverse in the domain $A=[0,\lambda]\times[0,\lambda]$

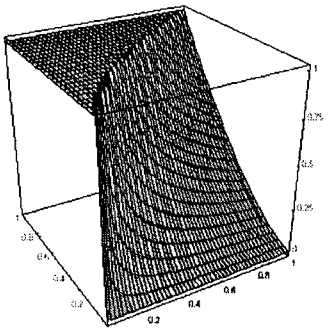


Figure 2.

Example 3.2

If $f(x) = 2(1\text{-}x)$ and $p=2/3$ then

$$f^{[-1]}(x)=\begin{cases} Min\left(1,1-\frac{x}{2}\right) & x\le 2 \\ 0 & \text{otherwise} \end{cases}$$

and

$$M(x,y)=f^{[-1]}\left[\frac{2}{3}(2-2x+2-2y)\right]=$$

$$Max\left(\frac{2x+2y-1}{3},0\right) \quad \text{(Figure 3)}$$

$$\hat{M}(x|y)=Min\left(1,\frac{1+3y-2x}{2}\right) \quad \text{for } x\in[0,1/2]$$

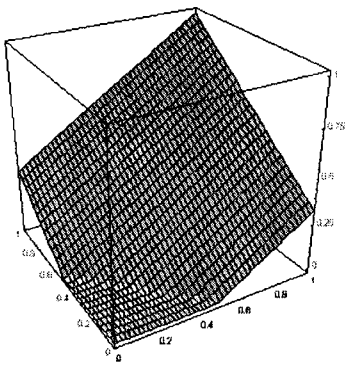


Figure 3

Remark 3.1

As it is shown in theorem 3.2, in the case of a non-strict operation with $p>1/2$, $\hat{M}(x|y)$ is not defined in the whole unit interval. In this case, we propose an extension of $\hat{M}(x|y)$ to the unit interval in the following way:

$$\hat{M}(x|y)=f^{[-1]}\left(\frac{f(y)}{p}-f(x)\right)$$

The function $\hat{M}(x|y)$ is well defined in the unit interval, it coincides with the residuation of M in the points where it exists and it has the value zero in the remaining points. Effectively, If $M(0,x)>y$ then

$$p(f(0)+f(x))<f(y)\Leftrightarrow \frac{f(y)}{p}-f(x)>f(0) \text{ and}$$

$$f^{[-1]}\left(\frac{f(y)}{p}-f(x)\right)=0$$

The generalised residuation for the example 3.2 is

$$\hat{M}(x|y)=Max\left(Min\left(1,\frac{1+3y-2x}{2}\right),0\right)$$

for $(x,y)\in I\times I$ (Figure 4)

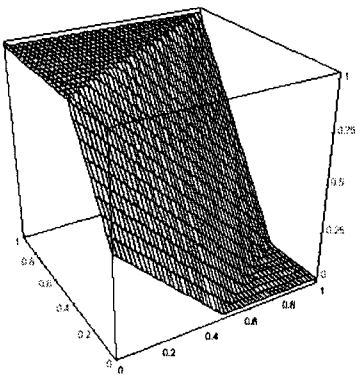


Figure 4: Generalised $\hat{M}(x|y)$

Let us denote by $\hat{\boldsymbol{M}}$ the subset of those operations of M that have residuation in the unit interval

Theorem 3.3

Let R be a M-transitive relation ($M\in\hat{\boldsymbol{M}}$) defined on a set X and $h_z = R(z,.)$ $(z\in X)$, then

$$R(x,y)\le \inf_{z\in X}\hat{M}(h_z(x)|h_z(y)) \qquad (3)$$

The preceding result indicates that any M-transitive relation ($M\in\hat{\boldsymbol{M}}$) has the relation

$$R^*=\inf_{z\in X} R_M\circ(h_z\times h_z)$$

as an upper bound, being R_M the residuation of M considered as a fuzzy relation in I.

On the other hand, it is worth noting that, in general, the relation R_M is not M-transitive (except for $p=1$). Therefore, given a fuzzy set $h\in[0,1]^X$, the fuzzy relation R defined by $R(x.y)=R_M(h(x)|h(y))$ $(p\neq 1)$ is not M-transitive and neither it is the relation (3) of theorem 3.3. Nevertheless, we have the following result

Theorem 3.4

If $M_f \in \hat{\boldsymbol{M}}$, then the relation R_{M_f} is T_f transitive

Remark 3.2

In the previous proof of theorem 3.4, the definition 2.3 of quasi-inverse has not been used. The reasoning is based on the definition proposed in the remark 3.2. Therefore, theorems 3.3 and 3.4 are also valid by using this second one that it only affects to the non-strict operations.
Summarising the two preceding theorems we have the following result

Theorem 3.5

If R is a M-transitive relation ($M \in \boldsymbol{M}$) on a set X then

$$R \le R^* = \inf_{z \in X} R_M \circ (h_z \times h_z), \qquad (4)$$

where $h_z(x)=R(z,x)$ and R^* is a T_f-transitive relation (T_f is the archimedean t-norm associate to M).
This result gives us a method to generate non-reflexive relations that are transitive with respect to an arcimedean t-norm.

Theorem 3.6

Given a family of fuzzy sets $\{\mu_j\}_{j \in J}$ of X and $M \in \hat{\boldsymbol{M}}$, the fuzzy relation

$$R^*(x,y) = \inf_{j \in J} R_M(\mu_j(x) | \mu_j(y))$$

is a T_f-transitive relation, being T_f the t-norm associated to M.

Remark 3.3

If $M = T_f$, (p=1), then R_M is a reflexive relation. R^* is a preorder and $R=R^*$ (4) since it is a particular case of the representation theorem for similarity relations [15].
Finally, by symmetrising the relation R^*, we can generate symmetric and transitive relations with respect to an archimedean t-norm, that are not necessarily reflexive.

Theorem 3.7

Given a family of fuzzy sets $\{\mu_j\}_{j \in J}$ of X and $M \in \hat{\boldsymbol{M}}$, the fuzzy relation

$$R^*(x,y) = \inf_{j \in J} R_M(Max(\mu_j(x), \mu_j(y)) | Min(\mu_j(x), \mu_j(y)))$$

is a symmetric and T_f-transitive relation and not necessarily reflexive.

References

[1] Aczél J. (1966), Lectures on Functional Equations and their Applications, Academic Press, New York-London.

[2] Alsina,C. (1984), Between Min and Max. Proc. FISAL-84, UIB 37-40.

[3] Alsina C., Trillas E.; Valverde L., (1981), On some Logical Connectives for Fuzzy Sets Theory. JMMAA. 93, 15-26.

[4] Alsina, C., Cubillo, S., Jacas, J., Trillas, E. (1992), On And in [0,1]. Proceedings IPMU 92, Palma de Mallorca.

[5] De Baets, B, Mesiar R.(1998), T-partitions. Fuzzy Sets & Systems vol. 97-2 , 211-223.

[6] Fodor J.C. (1992) Traces of fuzzy binary relations. Fuzzy Sets & Systems vol. 50-3 , 331-341.

[7] Jacas, J.; Recasens, J., (1993), Fuzzy T-transitive relations: eigenvectors and generators, Fuzzy Sets and Systems, vol. 72, 2, 147-154.

[8] Klement E. P., Mesiar R., Pap E. (1999). Quasi- and pseudo-inverses of monotone functions, and the construction of t-norms. Fuzzy Sets & Systems vol. 104-1 , 3-13.

[9] Poncaré, H. (1943), La Ciencia y la Hipótesis. Espasa-Calpe.

[10] Menger, K., Probabilistic theory of relations. Proc. Nat. Acad. Sci. USA, 37, (1951), 178-180.Poincaré, H., (1943), La Ciencia y la Hipótesis, Espasa-Calpe.

[11] Trillas, E., (1982), Assaig sobre les relacions d'indistinguibilitat. Actes del Primer Congrés Català de Lògica Matemàtica, Barcelona, 51-54.

[12] Trillas, E.; Alsina, C., (1992), Logic: going further from Tarski, Fuzzy Sets and Systems, North Holland..

[13] Trillas, E.; Valverde, L., (1984), On implication an indistinguishability in the setting of fuzzy logic, in: J. Kacprzyk and R.R. Yager Eds., Management Decision Support Systems Using Fuzzy Sets and Possibility Theory (Verlag, TUV, Rheinland).

[14] Trillas, E.; Valverde, L., (1984), An inquiry into indistinguishability operators in: A. Skala et alt. eds., Aspects of Vagueness (Reidel Dordrecht, 1984), 231-256.

[15] Valverde, L, (1985), On the structure of F-indistinguishability operators. Fuzzy Sets and Systems, 17, 313-328.

REPRESENTATIONS AND CONSTRUCTIONS OF STRONGLY LINEAR FUZZY ORDERINGS

Ulrich Bodenhofer
Dept. of Algebra, Stochastics, and Knowledge-Based Math. Systems
Johannes Kepler Universität
A-4040 Linz, Austria
ulrich@flll.uni-linz.ac.at

Summary

This paper is devoted to a class of fuzzy orderings which play a fundamental role in decision analysis and fuzzy control—strongly linear fuzzy (weak) orderings. First, we see that any relation of that kind can be decomposed into a crisp linear ordering and a fuzzy equivalence relation. As a consequence, a general representation theorem follows. Finally, a method for constructing strongly linear fuzzy orderings from pseudo-metrics is presented.

Keywords: Fuzzy ordering, fuzzy weak ordering, preference modeling.

1 INTRODUCTION

Fuzzy (weak) orderings have been studied intensively in the context of preference modeling [3, 5, 6]. Recent investigations [1, 2], moreover, have shown that they can also be integrated fruitfully in areas related to fuzzy systems.

Definition 1. Consider a t-norm T. A binary fuzzy relation $R : X^2 \to [0,1]$ is called

1. *reflexive* iff $\forall x \in X : \; R(x,x) = 1$,
2. *symmetric* iff $\forall x, y \in X : \; R(x,y) = R(y,x)$,
3. *T-transitive* iff

$$\forall x, y, z \in X : \; T\big(R(x,y), R(y,z)\big) \leq R(x,z),$$

4. *strongly linear* (strongly complete) iff

$$\forall x, y \in X : \; \max\big(R(x,y), R(y,x)\big) = 1.$$

Definition 2. Suppose that T is an arbitrary t-norm.

1. A reflexive and T-transitive fuzzy relation is called *fuzzy preordering* with respect to a t-norm T, short *T-preordering*.
2. Symmetric T-preorderings are called *fuzzy equivalence relation* with respect to T, short *T-equivalences*.
3. If a T-preordering is, in addition, strongly linear, it is called *fuzzy weak ordering* with respect to T, for brevity *weak T-ordering*.

It is known, at least if the underlying t-norm T is left-continuous, that T-preorderings and T-equivalences are, in some sense, generated by families of fuzzy subsets of X [8], while an analogous unique representation of fuzzy weak orderings is still missing. The purpose of this paper is (1) to close this gap and (2) to use some of the representation results to link fuzzy weak orderings with pseudo-metrics.

2 REPRESENTATIONS

As a first important result, we can show that any fuzzy weak ordering can be decomposed into a crisp ordering and a fuzzy equivalence relation. In order to be able to do so, we have to introduce the notion of compatibility between a crisp ordering and a fuzzy equivalence relation first.

Definition 3. Let $\lesssim$ be a crisp ordering on X and let E be a fuzzy equivalence relation on X. E is called *compatible with* $\lesssim$, if and only if the following implication holds for all $x, y, z \in X$:

$$x \lesssim y \lesssim z \Longrightarrow E(x,z) \leq \min\big(E(x,y), E(y,z)\big)$$

This property can be interpreted as follows: The two outer elements of a three-element chain are at most as similar as any two inner elements.

Theorem 4. *Consider a fuzzy relation L on a domain X. Then the following two statements are equivalent:*

(i) L is a weak T-ordering.

(ii) There exists a linear ordering $\lesssim$ and a T-equivalence E, which is compatible with $\lesssim$, such that L can be represented as follows:

$$L(x,y) = \begin{cases} 1 & \text{if } x \lesssim y \\ E(x,y) & \text{otherwise} \end{cases} \quad (1)$$

Proof. See [1]. □

The representation (1) implies that any weak T-ordering is uniquely characterized as a "direct fuzzification" of a crisp linear ordering $\lesssim$, where the fuzzy component can solely be attributed to a T-equivalence.

What follows next is a unique residuum-based characterization of fuzzy weak orderings. For the remaining section, assume that T denotes a left-continuous t-norm with the unique residual implication

$$\vec{T}(x,y) = \sup\{z \in [0,1] \mid T(x,z) \leq y\}$$

and its corresponding biimplication

$$\overleftrightarrow{T}(x,y) = \min\left(\vec{T}(x,y), \vec{T}(y,x)\right).$$

Valverde [8] has shown that any set of fuzzy subsets $(A_i)_{i\in I}$ generates a T-preordering by

$$L(x,y) = \inf_{i\in I} \vec{T}\left(\mu_{A_i}(x), \mu_{A_i}(y)\right). \quad (2)$$

The reverse is true as well, i.e. any T-preordering is generated by some family $(A_i)_{i\in I}$ in the above way.

An analogous representation holds for T-equivalences if we replace the implication $\vec{T}$ by the biimplication $\overleftrightarrow{T}$, i.e. a fuzzy relation E is a T-equivalence if and only if there exists a family of fuzzy subsets $(A_i)_{i\in I}$ such that the following holds [8]:

$$E(x,y) = \inf_{i\in I} \overleftrightarrow{T}\left(\mu_{A_i}(x), \mu_{A_i}(y)\right) \quad (3)$$

For fuzzy weak orderings, however, no such unique representation is available. Several researchers have considered fuzzy relations generated by single fuzzy sets [3, 5]

$$L(x,y) = \vec{T}\left(\mu_A(x), \mu_A(y)\right). \quad (4)$$

As follows from elementary properties of the implication $\vec{T}$, (4) always defines a weak T-ordering, while not every weak T-ordering can be represented in that way.

Theorem 4, however, enables us to formulate a general representation theorem for arbitrary fuzzy weak orderings with respect to left-continuous t-norms.

Theorem 5. *Provided that L is a binary fuzzy relation on a domain X and that T denotes a left-continuous t-norm, the following two statements are equivalent:*

(i) L is a fuzzy weak ordering with respect to T.

(ii) There exists a linear ordering $\lesssim$ on X and a family of fuzzy subsets $(A_i)_{i\in I}$ of X, the membership functions of which are non-decreasing with respect to $\lesssim$, such that the representation (2) holds.

Proof. (i)⇒(ii): First of all, Theorem 4 guarantees the existence of a linear ordering $\lesssim$, which is "below" L, i.e.

$$x \lesssim y \implies L(x,y) = 1. \quad (5)$$

Assuming $I = X$, we define

$$\mu_{x_0}(x) = L(x_0, x)$$

for all $x_0 \in X$. Since L is T-transitive and property (5) holds, $x \lesssim y$ implies

$$L(x_0,x) = T\big(L(x_0,x), \overbrace{L(x,y)}^{=1}\big) \leq L(x_0,y)$$

which equivalent to $\mu_{x_0}(x) \leq \mu_{x_0}(y)$ and we have shown that all membership functions μ_{x_0} are non-decreasing w.r.t. $\lesssim$. Then the representation (2) follows from reflexivity and T-transitivity as in the proof of [8, Theorem 4.1].

(ii)⇒(i): Analogous to the proof of [8, Theorem 4.1], where strong linearity immediately follows from the monotonicity of all μ_{A_i}. □

It is worth to mention that, if we start from a given weak T-ordering L, neither the ordering $\lesssim$ nor the family $(A_i)_{i\in I}$ need to be unique.

Finally, we can adapt the representation (3) to the case of compatibility with a given crisp linear ordering.

Theorem 6. *Consider a crisp linear ordering $\lesssim$ of the domain X and a fuzzy relation $E : X^2 \to [0,1]$. Then the following two statements are equivalent:*

(i) E is a T-equivalence which is compatible with $\lesssim$.

(ii) There exists a family of fuzzy subsets $(A_i)_{i\in I}$, the membership functions of which are non-decreasing with respect to $\lesssim$, such that the representation (3) holds.

Proof. (i)⇒(ii): First of all, by (1), we can define a weak T-ordering L. It is easy to see that E is exactly the symmetric kernel of L:

$$E(x,y) = \min\left(L(x,y), L(y,x)\right)$$

Since, for the relation L, Theorem 5 is applicable, we obtain

$$\begin{aligned} E(x,y)\min\big(L(x,y),L(y,x)\big) &= \min\big(\inf_{i\in I}\vec{T}(\mu_{A_i}(x),\mu_{A_i}(y)),\\ &\qquad \inf_{i\in I}\vec{T}(\mu_{A_i}(y),\mu_{A_i}(x))\big)\\ &= \inf_{i\in I}\min\big(\vec{T}(\mu_{A_i}(x),\mu_{A_i}(y)),\\ &\qquad \vec{T}(\mu_{A_i}(y),\mu_{A_i}(x))\big)\\ &= \inf_{i\in I}\overleftrightarrow{T}(\mu_{A_i}(x),\mu_{A_i}(y)). \end{aligned}$$

(ii)⇒(i): Similar to [8, Theorem 4.2], where the compatibility easily follows from the monotonicity of all μ_{A_i}. □

3 CONSTRUCTIONS

In this section, we will discuss in which way strongly linear fuzzy orderings can be constructed provided that some linear ordering of the given universe is known. Of course, Theorem 5 provides the most flexible way to construct such orderings. Alternatively, we will now consider how, by means of Theorem 4, strongly linear fuzzy orderings can be constructed from pseudo-metrics, at least if the considered t-norm T is continuous and Archimedean, i.e.

$$\forall x \in (0,1):\ T(x,x) < x.$$

We will utilize the well-known correspondence between pseudo-metrics and T-equivalences (e.g. [4]) and extend this concept in order to integrate compatibility with a given crisp linear ordering.

Before turning to the core, let us briefly recall some important preliminaries.

Theorem 7. *A function $T : [0,1]^2 \to [0,1]$ is a continuous Archimedean t-norm if and only if there exists a continuous, strictly decreasing function $f : [0,1] \to [0,\infty]$ with $f(1) = 0$ called* additive generator *such that, for all $x, y \in [0,1]$, the following holds:*

$$T(x,y) = f^{-1}\big(\min(f(x)+f(y), f(0))\big)$$

The generator f is uniquely determined up to a positive multiplicative constant.

Proof. See e.g. [7]. □

Definition 8. A mapping $d : X^2 \to [0,\infty]$ is called *pseudo-metric* on X if and only if the following axioms hold:

1. *Homogeneity:* $\forall x \in X:\ d(x,x) = 0$
2. *Symmetry:* $\forall x, y \in X:\ d(x,y) = d(y,x)$
3. *Triangle inequality:*

$$\forall x, y, z \in X:\ d(x,z) \le d(x,y) + d(y,z)$$

Theorem 9. *Assume that T is an Archimedean t-norm with an additive generator f.*

1. *For any pseudo-metric $d : X^2 \to [0,\infty]$, the mapping*

$$E_d(x,y) = f^{-1}\big(\min(d(x,y), f(0))\big) \qquad (6)$$

defines a T-equivalence on X.

2. *Provided that E is a T-equivalence on X, we can define a pseudo-metric d_E as*

$$d_E(x,y) = f(E(x,y)). \qquad (7)$$

Proof. See [4]. □

The following lemma induces a dual notion of compatibility between pseudo-metrics and orderings and characterizes its correspondence with the compatibility between fuzzy equivalence relations and orderings.

Lemma 10. *Let T be a continuous Archimedean t-norm with an additive generator f and let $\lesssim$ be an ordering of the domain X.*

1. *If a pseudo-metric d on X has the property*

$$\forall x, y, z \in X:\ x \lesssim y \lesssim z \Longrightarrow d(x,z) \ge \max(d(x,y), d(y,z)), \qquad (8)$$

then its induced fuzzy equivalence relation E_d, defined as in (6), is compatible with $\lesssim$.

2. *If a fuzzy equivalence relation E is compatible with $\lesssim$, its induced pseudo-metric d_E, defined as in (7), fulfills property (8).*

Proof. Follows directly from the fact that the additive generator f and its inverse are non-increasing function. □

Moreover, we obtain that all desired properties remain even after transformation with some monotonic (not necessarily bijective) function.

Lemma 11. *Consider a pseudo-metric $d : X^2 \to [0,1]$, which is compatible with a linear ordering $\lesssim$, and a non-decreasing function $\varphi : X \to X$. Then the mapping*

$$d'(x,y) = d(\varphi(x), \varphi(y))$$

also defines a pseudo-metric on X which is compatible with $\lesssim$.

Proof. Trivial. □

Putting all the above results together, we obtain the following construction theorem.

Theorem 12. *Assume that we are given a crisp linear ordering $\lesssim$ of a space X, a pseudo-metric d, which is compatible with $\lesssim$ in the sense of (8), a continuous Archimedean t-norm T with an additive generator f, and a non-decreasing function $\varphi : X \to X$. Then*

$$L_{\varphi,d}(x,y) = \begin{cases} 1 & \text{if } x \lesssim y \\ E_{\varphi,d}(x,y) & \text{otherwise} \end{cases}$$

defines a strongly linear fuzzy ordering with respect to T and

$$E_{\varphi,d}(x,y) = f^{-1}\big(\min(d(\varphi(x),\varphi(y)), f(0))\big).$$

Proof. Immediate consequence of Theorem 4 and the above results. □

Example 13. It is easy to see that $d(x,y) = |x-y|$ is a (pseudo-)metric on the real numbers which is compatible with their natural ordering $\leq$. It is, moreover, straightforward to check that $1-x$ is a self-inverse additive generator of the Łukasiewicz t-norm

$$T_{\mathbf{L}}(x,y) = \max(x+y-1, 0)$$

and that e^{-x} is an additive generator of the product t-norm $T_{\mathbf{P}}$, whose inverse is $-\ln x$. Hence, for any non-decreasing function $\varphi : \mathbb{R} \to \mathbb{R}$, we obtain that

$$\begin{aligned} E'_{\varphi}(x,y) &= 1 - (\min(|\varphi(x) - \varphi(y)|, 1)) \\ &= \max(1 - |\varphi(x) - \varphi(y)|, 0) \end{aligned}$$

is a $T_{\mathbf{L}}$-equivalence which is compatible with $\leq$ and that

$$E''_{\varphi}(x,y) = e^{-\min(|\varphi(x)-\varphi(y)|,\infty)} = e^{-|\varphi(x)-\varphi(y)|}$$

is a $T_{\mathbf{P}}$-equivalence which is compatible with $\leq$. Then, based on Theorem 12, we can define strongly linear fuzzy orderings on the real numbers, which fuzzify the linear ordering $\leq$.

Not surprisingly, we can also revert this construction: Starting from an arbitrary weak T-ordering L, by means of Theorem 4, we obtain a linear ordering $\lesssim$ and a T-equivalence E compatible with $\lesssim$. Then, by Theorem 9 and Lemma 10, we can construct a pseudo-metric which is compatible with $\lesssim$ as well.

4 CONCLUDING REMARKS

We have seen in which way fuzzy weak orderings are characterized. On the one hand, they can be considered as combinations (unions) of crisp linear orderings and fuzzy equivalence relations. On the other hand, they are uniquely characterized as (possibly infinite) min-intersections of fuzzy weak orderings induced by fuzzy sets with monotonic membership functions. Perhaps the most important discovery is that, even if intermediate degrees of preference are allowed, there is an underlying crisp ordering.

In the second part—soundly extending the well-known connection between pseudo-metrics and fuzzy equivalence relations—we have considered the correspondence between fuzzy weak orderings and pseudo-metrics.

The next step is to study the consequences of these results in decision making, in particular concerning aggregation of preferences.

Acknowledgements

The author gratefully appreciates partial support of the Austrian "Fonds zur Förderung der wissenschaftlichen Forschung" (project P-12900-TEC) and the CEEPUS network SK-42.

References

[1] U. Bodenhofer (1998). *A Similarity-Based Generalization of Fuzzy Orderings.* PhD thesis, Johannes Kepler Universität Linz.

[2] U. Bodenhofer (1999). The construction of ordering-based modifiers. In G. Brewka, R. Der, S. Gottwald, and A. Schierwagen, editors, *Fuzzy-Neuro Systems '99*, pages 55–62. Leipziger Universitätsverlag.

[3] B. De Baets, J. Fodor, and E. E. Kerre (1999). Gödel representable fuzzy weak orders. *Internat. J. Uncertain. Fuzziness Knowledge-Based Systems*, 7(2):135–154.

[4] B. De Baets and R. Mesiar (1997). Pseudo-metrics and T-equivalences. *J. Fuzzy Math.*, 5(2):471–481.

[5] J. Fodor and S. V. Ovchinnikov (1995). On aggregation of t-transitive fuzzy binary relations. *Fuzzy Sets and Systems*, 72:135–145.

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

[7] B. Schweizer and A. Sklar (1983). *Probabilistic Metric Spaces.* North-Holland, Amsterdam.

[8] L. Valverde (1985). On the structure of F-indistinguishability operators. *Fuzzy Sets and Systems*, 17(3):313–328.

Divergence Measures and Aggregation Operations

Miranda, P.
Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo-Spain
est010@pinon.ccu.uniovi.es

Torres, E.
Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo-Spain
torres@etsiig.uniovi.es

Gil, P.
Dpto. de Estadística, I. O. y D. M.
Universidad de Oviedo-Spain
pedro@pinon.ccu.uniovi.es

Summary

In this paper we deal with the relations between divergence measures between two fuzzy sets and aggregation operations while having a finite universal set. We will prove that aggregation operations can be related with divergence measures. We feel that this can be useful in order to study some properties of these last measures. Finally, we find some relations between divergence measures and measures of fuzziness.

Keywords: Divergence measures, Aggregation Operations, Measures of Fuzziness.

1 Introduction and basic concepts

In the following, divergence measures are denoted as D; the universal set is denoted by Ω; fuzzy subsets are denoted by A, B and so on; n denotes the cardinality of Ω; elements of Ω are denoted by x_i. An aggregation operation is denoted by s and a measure of fuzziness is denoted by f.

Although divergence measures and measures of fuzziness can attain the infinite value we will not consider this case (see remark 2 above).

Definition 1 *Let Ω be the universal set and let $\widetilde{\mathcal{P}}(\Omega)$ be the set of all fuzzy subsets. A mapping $D : \widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \mapsto \mathbb{R}$ is said to be a* **divergence measure between fuzzy subsets** *[7] if and only if the three following axioms hold for any $A, B, C \in \widetilde{\mathcal{P}}(\Omega)$:*

1. $D(A, B) = D(B, A)$.
2. $D(A, A) = 0$.
3. $\max\{D(A \cup C, B \cup C), D(A \cap C, B \cap C)\} \leq D(A, B)$, *in which we have considered the usual definitions based on* max *and* min.

The most common divergence measures are those called local divergence measures. For these measures, each coordinate is independent from the others and they are all equally important. Formally, we have the following definition:

Definition 2 *Let D be a divergence measure over a finite universal set Ω. We say that D is* **local** *if and only if there exists a function $h : [0,1] \times [0,1] \mapsto \mathbb{R}$ verifying*

$$D(A, B) - D(A \cup \Omega^i, B \cup \Omega^i) = h(A(x_i), B(x_i)),$$

$$\forall A, B \in \widetilde{\mathcal{P}}(\Omega), \forall x_i \in \Omega \qquad (1)$$

where

$$\Omega^i(x_j) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

For these special divergence measures the following can be proved [7]:

Proposition 1 *If we have a finite universal set Ω, then D is a local divergence measure if and only if there exists a mapping $h : [0,1] \times [0,1] \mapsto \mathbb{R}$ such that*

$$D(A, B) = \sum_{i=1}^{n} h(A(x_i), B(x_i)). \qquad (2)$$

verifying the following conditions:

- $h(x, y) = h(y, x), \forall x, y \in [0, 1]$;
- $h(x, x) = 0, \forall x \in [0, 1]$;
- $h(\cdot, y)$ *is a non-increasing function on [0,y] and non-decreasing on [y,1].*

The aggregation operations on fuzzy sets are operations by which several fuzzy sets are combined to produce a single set. In particular, fuzzy unions and fuzzy intersections are special cases of aggregation operations. There are a lot of definitions of aggregation operations, depending on the problem to be solved and even on the author. We will follow the definition given by Klir&Folger in [5].

Definition 3 *A* **aggregation operation** *is defined by a function* $s : [0,1]^n \mapsto [0,1]$ *verifying*

- $s(\vec{0}) = 0, s(\vec{1}) = 1$ *(boundary conditions).*
- *If* $\vec{a} \leq \vec{b} \Rightarrow s(\vec{a}) \leq s(\vec{b})$ *where* $\vec{a} \leq \vec{b} \Leftrightarrow a_i \leq b_i$ *(monotonicity).*

2 Relationship between divergence measures and aggregation operations

Theorem 1 *Let* $h_1, ..., h_n$ *be* n *mappings,* $h_i : [0,1] \times [0,1] \mapsto [0,1]$, $\forall i$ *verifying:*

- $h_i(x,x) = 0$, $\forall x \in [0,1]$.
- $h_i(x,y) = h_i(y,x)$, $\forall x, y \in [0,1]$.
- $h_i(1,0) = h_i(0,1) = 1$.
- $h_i(\cdot, y)$ *is a non-increasing function over* $[0,y]$ *and non-decreasing over* $[y,1]$.

Let s *be an aggregation operation. Then,*

$$D(A,B) = ks(h_1(A(x_1), B(x_1)), ..., h_n(A(x_n), B(x_n))) \quad (3)$$

with $k > 0$ *is a divergence measure.*

Proof. Let s be an aggregation operation and let us see that D is a divergence measure.

- $D(A,B) = D(B,A)$ is trivial using condition 2 over h_i.
- $D(A,A) = ks(\vec{0}) = 0$ using condition 1 over h_i.
- $D(A \cup C, B \cup C) = ks(h_1((A \cup C)(x_1), (B \cup C)(x_1)), ..., h_n((A \cup C)(x_n), (B \cup C)(x_n)))$, $D(A,B) = ks(h_1(A(x_1), B(x_1), ..., h_n(A(x_n), B(x_n)))$. Let us see that $h_i((A \cup C)(x_i), (B \cup C)(x_i)) \leq h_i(A(x_i), B(x_i))$. We have three cases:
  1. If $C(x_i) \geq A(x_i)$, $C(x_i) \geq B(x_i)$, then $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(C(x_i), C(x_i)) = 0 \leq h_i(A(x_i), B(x_i))$.
  2. If $A(x_i) \geq C(x_i) \geq B(x_i)$ or $B(x_i) \geq C(x_i) \geq A(x_i)$, then (considering the first possibility) $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(A(x_i), C(x_i)) \leq h_i(A(x_i), B(x_i))$ because $h_i(x,z) \geq h_i(x,y)$ if $x \geq y \geq z$ (using condition 4 over h_i).
  3. If $A(x_i) \geq C(x_i)$, $B(x_i) \geq C(x_i)$, then $h_i((A \cup C)(x_i), (B \cup C)(x_i)) = h_i(A(x_i), B(x_i))$.

  The same holds for the intersection.

Then D is a divergence measure ■

Let us give some remarks:

1. Condition 3 over h_i is needed in order to s can be applied. We could have considered $h_i(x,y) \leq 1$.
2. Divergence measures D obtained in the theorem verify that: If $h_i(A(x_i), B(x_i)) = h_i(C(x_i), D(x_i)), \forall i$, then $D(A,B) = D(C,D)$ as s is applied over the same vectors.
3. k is needed because divergence measures can reach values greater than 1.

Note that the conditions over h_i are the same as those for local divergence measures. Thus, we have the following

Corollary 1 *If we take* $h_i = g, \forall i$ *and we take the arithmetic mean as aggregation operation we obtain a local measure.*

In fact, note that all local divergence measures can be generated like this, taking $k = n$.

Theorem 2 *Let* D *be a divergence measure. Then,*

$$s(\vec{p}) = \frac{D(A,\emptyset)}{D(\Omega,\emptyset)} \quad (4)$$

with $A(x_i) = p_i$ *is an aggregation operation.*

Proof. Let us see that the axioms for aggregation operation hold:

- $s(\vec{1}) = \frac{D(\Omega,\emptyset)}{D(\Omega,\emptyset)}$ and thus $s(\vec{1}) = 1$. Now, $s(\vec{0}) = \frac{D(\emptyset,\emptyset)}{D(\Omega,\emptyset)}$ and then $s(\vec{0}) = 0$.
- Let $\vec{p} \leq \vec{q}$. Then, if we define $A(x_i) = p_i$, $B(x_i) = q_i$, it suffices to show that $D(B,\emptyset) \geq D(A,\emptyset)$. But this is obvious because $B(x_i) \geq A(x_i)$ and thus, using the third condition of divergence measures we obtain

$$D(B,\emptyset) \geq D(B \cap A, \emptyset \cap A) = D(A,\emptyset).$$

Then s is an aggregation operation ■

Remark 1 *Maybe it is not possible to find s given D and h_i. To see this, consider $|\Omega| = 1, D(A,B) = |A(x) - B(x)|$ and*

$$h(x,y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

Then, D can take infinite values and h can take only two. Thus, it is not possible to find an aggregation operation relating them because there are different pairs of subsets $(A,B),(C,D)$ such that $h_i(A(x_i),B(x_i)) = h_i(C(x_i),D(x_i)), \forall i$ but $D(A,B) \neq D(C,D)$.

Theorem 1 and a generalisation of theorem 2 can be put together as:

Theorem 3 *Let $h_1,...,h_n$ be n mappings, $h_i : [0,1] \times [0,1] \mapsto [0,1]$ verifying:*

- *$h_i(x,x) = 0, \forall x \in [0,1]$.*
- *$h_i(1,0) = h_i(0,1) = 1$.*
- *$h_i(x,y) = h_i(y,x), \forall x,y \in [0,1]$.*
- *$h_i(\cdot,y)$ is a non-increasing function over $[0,y]$ and non-decreasing over $[y,1]$.*

and let D be a set function verifying that if $h_i(A(x_i),B(x_i)) = h_i(C(x_i),D(x_i)), \forall i$, then $D(A,B) = D(C,D)$. Then, D is a divergence measure if and only if it can be put as

$$D(A,B) = ks(h_1(A(x_1),B(x_1)),...,h_n(A(x_n),B(x_n))) \quad (5)$$

where $k > 0$ and s is an aggregation operator.

Proof.⇒ Let D and h_i be as in the theorem. Let us see that we can define s an aggregation operation. We define $s(\vec{p}) = \frac{D(A,\emptyset)}{D(\Omega,\emptyset)}$ with $A(x_i) = sup_{h_i(0,x')\leq p_i} x'$.

- $s(\vec{1}) = \frac{D(\Omega,\emptyset)}{D(\Omega,\emptyset)}$ and thus $s(\vec{1}) = 1$. Now, $s(\vec{0}) = \frac{D(\emptyset,\emptyset)}{D(\Omega,\emptyset)}$ and then $s(\vec{0}) = 0$.
- Let $\vec{p} \leq \vec{q}$. Then, if we define $A(x_i) = sup_{h_i(0,x')\leq p_i} x', B(x_i) = sup_{h_i(0,x')\leq q_i} x'$, it suffices to show that $D(B,\emptyset) \geq D(A,\emptyset)$. But this is obvious because $B(x_i) \geq A(x_i)$ and thus, using the third condition of divergence measures we obtain

$$D(B,\emptyset) \geq D(B \cap A, \emptyset \cap A) = D(A,\emptyset).$$

⇐ Theorem 1 ■

Remark 2 *Divergence measures can assume the ∞ value. Consider, for example $D(A,B) = \sum_{x\in\Omega} |A(x) - B(X)|$ when $|\Omega| = \infty$. However, in the finite case it is not really a useful measure.*

3 Relationship between divergence measures and measures of fuzziness

Measures of fuzziness appear in the seventies to measure the vagueness or fuzziness of a fuzzy set. They have been extensively studied [2], [10].

Definition 4 *A* **measure of fuzziness***[6] is a mapping*

$$f : \widetilde{\mathcal{P}}(\Omega) \mapsto \mathbb{R}^+$$

verifying

- *$f(A) = 0$ if and only if A is a crisp subset.*
- *If $A \preceq B$ then $f(A) \leq f(B)$ where $A \preceq B$ denotes that A is sharper (less fuzzy) than or as fuzzy as B.*
- *$f(A)$ assumes the maximum value if and only if A is maximally fuzzy. We will denote the maximally fuzzy subset by M.*

We will consider $A \preceq B \Leftrightarrow |A(x) - A^c(x)| \leq |B(x) - B^c(x)|, \forall x \in \Omega$ with c a fuzzy complement.

Now, the following can be proved:

Theorem 4 *Consider the classical fuzzy union and fuzzy intersection proposed by Zadeh. Let c be an involutive fuzzy complement. Let f be a mapping over fuzzy subsets such that $f(A) = f(A^c)$. Then, f is a measure of fuzziness if and only if f can be put as*

$$f(A) = k(D(Z,Z^c) - D(A,A^c)), k > 0 \quad (6)$$

with Z a crisp subset and D a normalized divergence measure verifying

- *$D(A,A^c) = 1 \Leftrightarrow A$ crisp.*
- *$D(A,B) = 0 \Leftrightarrow A = B$.*
- *$D(A,B)$ only depends on $|A(x) - B(x)|, \forall x \in \Omega$, i.e., if A,B,C,D are such that $|A(x) - B(x)| = |C(x) - D(x)|, \forall x \in \Omega$, then $D(A,B) = D(C,D)$.*

Proof.⇒Let f be a measure of fuzziness. Let us see that given $A,B, \exists C | D(C,C^c) = D(A,B)$. As c is a continuous complement (c is involutive), the mapping $c - c^2 = c - Id$ is also a continuous function. Now, $c(0) - 0 = 1, c(1) - 1 = -1$. Then, for each $x \in \Omega$ there exists v such that $c(v) - v = |A(x) - B(x)|$. This

value of v is unique: Suppose that there exists another value $v' > v$ verifying $c(v') - v' = |A(x) - B(x)|$. Then, as c is a non-increasing function, $c(v') \leq c(v)$ and thus $c(v') - v' < c(v) - v$ contradicting our hypothesis. We define $C_{A,B}(x) = v$. Using our hypothesis 3 over D we obtain that $D(C_{A,B}, C^c_{A,B}) = D(A, B)$.

We define $D(C_{A,B}, C^c_{A,B}) = \frac{f(M) - f(C_{A,B})}{k}$. By hypothesis 3, D is well-defined. Now, let us see that D is a divergence measure.

- As $|A(x) - A(x)| = 0$, then $C_{A,A}(x) = e$, the equilibrium point of c (this point always exists and is unique for continuous complements). Thus, $C_{A,A} = M$ and we obtain $D(A, A) = 0$.
- $D(A, B) = D(B, A)$ is trivial.
- It is trivial. Just note that $C_{A,B}$ is sharper than $C_{A \cup D, B \cup D}$ and thus $f(C_{A \cup D, B \cup D}) \geq f(C_{A,B})$ and thus $D(A, B) \geq D(A \cup D, B \cup D)$. The same holds for the intersection.

Then D is a divergence measure. It is easy to see that it verifies the conditions of the theorem.

Note also that for a crisp subset Z, given Z, Z^c, $C_{Z,Z^c} = \emptyset$ and then $D(Z, Z^c) = f(M)$. Thus, $f(A) = D(Z, Z^c) - D(A, A^c)$.

$\Leftarrow$Let D be as in the theorem and let us see that f is a measure of fuzziness.

- If A crisp, $f(A) = 0$ by hypothesis 1 and these are the only subsets verifying this condition.
- Let $A \preceq B$, then $D(A, A^c) = D(A', A'^c)$ with

$$A'(x) = \begin{cases} A(x) & if \quad A(x) \leq e \\ A^c(x) & if \quad A(x) > e \end{cases}$$

  Now, we have $D(A', A'^c) \geq D(A' \cup B', A'^c \cup B') = D(B', A'^c) \geq D(B' \cap B'^c, A'^c \cap B'^c) = D(B', B'^c) = D(B, B^c)$, and thus $f(A) \leq f(B)$.
- If M is the maximally fuzzy subset, we have $M = M^c$ and thus $D(M, M^c) = 0$. Then, f takes its maximum in M. Note that this is the only maximum by hypothesis 2 ■

Let us see some remarks:

1. If we consider general not-normalized divergence measures k can be removed.
2. We could make the same process as in the first part of the paper with general operators h_i instead of the absolute value. Then, we would have obtained similar results.
3. Note that all hypothesis are verified by the usual measures of fuzziness and divergence measures. In fact, $f(A) = f(A^c)$ is considered as an axiom by many authors [3].
4. This result was proved by Montes *et al.* in [8] for local divergence measures. Our result is an extension of this one: To prove that we have really an extension we have to find a non-local divergence measure verifying the conditions of the theorem. It is easy to see that for $\Omega = \{x_1, x_2\}$,

   $$D(A, B) = (1 + |A(x_1) - B(x_1)|)(1 + |A(x_2) - B(x_2)|) - 1$$

   is a divergence measure in these conditions.

Acknowledgements

This research has been supported in part by DGES grant number PB97-1286.

References

[1] B. De Baets & R. Mesiar (1998). T-partitions. *Fuzzy Sets and Systems 97* pp 213-223

[2] P. Benvenutti, D. Vivona & M. Divari (1998). Fuzziness Measures Via Pseudometrics on the Fuzzy Sets Space. *IPMU'98, Paris* pp. 1036-1039

[3] D. Dubois & H. Prade (1980). Fuzzy Sets and Systems. Theory and Applications. *Academic Press.*

[4] A. Kaufmann (1975). Introduction to the Theory of Fuzzy Subsets. *Academic Press, New York*

[5] G. Klir & T. Folger (1988). Fuzzy Sets, Uncertainty and Information. *Prentice Hall P T R*

[6] A. De Luca & E. Termini (1972). A definition of a nonprobabilistic entropy in the setting of fuzzy sets theory. *Information and Control, 20* pp. 301-312.

[7] S. Montes (1998) Particiones y medidas de divergencia en modelos difusos.*Ph. Thesis. University of Oviedo.*

[8] S. Montes, I. Couso& C. Bertoluzza (1999). Fuzziness Measures from Divergences. *EUROFUSE-SIC'99, Budapest* pp. 365-370

[9] S. Montes, P. Gil & C. Bertoluzza (1998). Divergence measure between Fuzzy sets and fuzziness. *IPMU'98, Paris* pp 1030-1035.

[10] E. Trillas & T. Riera. Entropies in finite fuzzy sets. *Inf. Sci. 15* pp 159-168.

DIVERGENCE MEASURES BASED ON T_∞

Torres, E.
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo-SPAIN
torres@etsiig.uniovi.es

Miranda, P.
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo-SPAIN
est10@pinon.ccu.uniovi.es

Gil, P.
Dpto. de Estadística, I.O. y D.M.
Universidad de Oviedo-SPAIN
pedro@pinon.ccu.uniovi.es

Summary

In this paper we deal with divergence measures between two fuzzy sets when we have a partition of the data. We characterize this divergence when all data are equally important and we show the relation between divergence measures, fuzzy measures and pseudo metrics.

Palabras Clave: Fuzzy Partition, Divergence Measure, Fuzzy Measure.

1 INTRODUCTION AND BASIC CONCEPTS

The primary objective of clustering techniques is to partition a given data set into homogeneous clusters, and a practical problem that clustering methods must address is how to measure the similarity. In this sense, divergence measures [1], [2], [3], [4], [7], are used in order to quantify the difference between two fuzzy subsets of a universal set Ω. These measures try to maintain the properties of classical divergence measures between two probability distributions which appear in Information Theory [6],[9]. We can also use divergence measures to quantify the separation among clusters on a fuzzy partition, and our paper is a first step in the analysis of these results using divergence measures.

In the following, divergence measures are denoted as D, fuzzy subsets are denoted by A, B, Ω is the universal finite set, $|\Omega| = n$, and S_∞ fuzzy union and T_∞ intersection are considered, i.e. $(A \cup B)(x) = \min\{A(x) + B(x), 1\}$ and $(A \cap B)(x) = \max\{A(x) + B(x) - 1, 0\}$. We use these t-norm and t-conorm due to their relation with fuzzy partition [5].

Let us now give some definitions:

Definition 1 *Let Ω be the universal set and let $\widetilde{\mathcal{P}}(\Omega)$ be the set of all fuzzy subsets. An application $D : \widetilde{\mathcal{P}}(\Omega) \times \widetilde{\mathcal{P}}(\Omega) \to \overline{\mathbb{R}}$ is said to be a* **divergence measure between fuzzy subsets** *if and only if the three following axioms hold:*

d1. $D(A, B) = D(B, A)$.

d2. $D(A, A) = 0$.

d3. $\max\{D(A \cup C, B \cup C), D(A \cap C, B \cap C)\} \leq D(A, B)$, $\forall C \in \widetilde{\mathcal{P}}(\Omega)$.

for any $A, B, C \in \widetilde{\mathcal{P}}(\Omega)$.

Definition 2 *Let D be a divergence measure. We say that D is* **local** *if and only if there exists a function $h : [0,1] \times [0,1] \to \overline{\mathbb{R}}$ verifying*

$$D(A, B) - D(A \cup \Omega^i, B \cup \Omega^i) = h(A(x_i), B(x_i)), \quad \forall A, B \in \widetilde{\mathcal{P}}(\Omega), \quad \forall x_i \in \Omega$$

where

$$\Omega^i(x) = \begin{cases} 1 & \text{if } x = x_i \\ 0 & \text{otherwise} \end{cases}$$

In fact, a local divergence measure says that each coordinate is independent of the others and that they are all equally important.

We can easily extend the definition of fuzzy measure to fuzzy subsets as:

Definition 3 *A* **fuzzy measure** *on $\widetilde{\mathcal{P}}(\Omega)$ is a set function $\mu : \widetilde{\mathcal{P}}(\Omega) \to [0, 1]$ satisfying the following axioms:*

1. $\mu(\emptyset) = 0$, $\mu(\Omega) = 1$.
2. $A \subset B \subset \Omega$ *implies* $\mu(A) \leq \mu(B)$.

Till the end, we will consider this definition.

2 DIVERGENCE MEASURES BASED ON T_∞

Proposition 1 *Let D be a divergence measure. Then $D(A,B) \geq 0$, $\forall A,\ B \in \widetilde{\mathcal{P}}(\Omega)$.*

Proof. From

$$D(A,B) \geq D(A \cap \emptyset, B \cap \emptyset) = D(\emptyset,\emptyset) = 0$$

■

Proposition 2 *Let A, B, C be fuzzy sets such that $A \subseteq B \subseteq C$. Then $D(A,B) \leq D(A,C)$ and $D(B,C) \leq D(A,C)$.*

Proof. To show that $D(A,B) \leq D(A,C)$, let H_1, H_2 be fuzzy sets such that $H_1(x) = (1 - C(x)) + (B(x) - A(x))$ and $H_2(x) = A(x)$, $\forall x \in \Omega$. H_1 is well defined since $A(x) \leq B(x) \leq C(x) \leq 1$.

$$\begin{aligned}
&(A \cap H_1)(x) = \\
&\quad = \max\{\ A(x) + (1 - C(x)) + \\
&\qquad\quad +(B(x) - A(x)) - 1, 0\} = 0 \\
&((A \cap H_1) \cup H_2)(x) = \\
&\qquad\quad = \min\{0 + A(x), 1\} = A(x)
\end{aligned}$$

$$\begin{aligned}
&(C \cap H_1)(x) = \\
&\quad = \max\{\ C(x) + (1 - C(x)) + \\
&\qquad\quad +(B(x) - A(x)) - 1, 0\} \\
&\quad = \quad B(x) - A(x) \\
&((C \cap H_1) \cup H_2)(x) = \\
&\quad = \min\{\ (B(x) - A(x)) + A(x), 1\} = B(x)
\end{aligned}$$

Thus

$$\begin{aligned}
&D(A,C) \geq \\
&\quad \geq D(A \cap H_1, C \cap H_1) \geq \\
&\quad \geq D((A \cap H_1) \cup H_2, (C \cap H_1) \cup H_2) = D(A,B)
\end{aligned}$$

To show that $D(B,C) \leq D(A,C)$, let be $H_3(x) = 1 - B(x)$, $H_4(x) = B(x)$ then

$$\begin{aligned}
&((A \cap H_3) \cup H_4)(x) = \\
&\quad = \min\{\ \max\{A(x) + (1 - B(x)) - 1, 0\} + \\
&\qquad\quad +B(x), 1\} \\
&\quad = \min\{\ 0 + B(x), 1\} = B(x)
\end{aligned}$$

$$\begin{aligned}
&((C \cap H_3) \cup H_4)(x) = \\
&\quad = \min\{\ \max\{C(x) + (1 - B(x)) - 1, 0\} + \\
&\qquad\quad +B(x), 1\} \\
&\quad = \min\{\ C(x) - B(x) + B(x), 1\} = C(x)
\end{aligned}$$

Then $D(B,C) \leq D(A,C)$. ■

Corollary 1 *If $A \subseteq B \subseteq C \subseteq H$ then $D(B,C) \leq D(A,H)$.*

Theorem 1 *If we have a finite universal set Ω, then D is a local divergence measure if and only if there exists an application $g : [0,1] \to R$ such that*

$$D(A,B) = \sum_{i=1}^{n} g(|A(x_i) - B(x_i)|),$$

verifying the following conditions:

- $g(0) = 0$;
- $g(\cdot)$ *is a monotonous non decreasing function in* $[0,1]$.

Proof. Let be $A_m(x) = (A \cup (\cup_{i=1}^{m}\{x_i\}))(x)$, and $B_m(x) = (B \cup (\cup_{i=1}^{m}\{x_i\}))(x)$, $m = 0, \ldots, n$, $\forall x \in \Omega$, thus $A_{m+1} = A_m \cup \{x_{m+1}\}$, and $A_n = \Omega = B_n$. If D is local, then

$$\begin{aligned}
&D(A_m, B_m) - D(A_m \cup \{x_{m+1}\}, B_m \cup \{x_{m+1}\}) = \\
&\quad = h(A(x_{m+1}), B(x_{m+1})),\ m = 0, \ldots, n-1.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\sum_{i=0}^{n-1} D(A_m, B_m) - \\
&\qquad - \sum_{i=0}^{n-1} D(A_m \cup \{x_{m+1}\}, B_m \cup \{x_{m+1}\}) = \\
&\qquad = D(A,B) - D(\Omega,\Omega) = D(A,B) \\
&\qquad = \sum_{i=1}^{n} h(A(x_i), B(x_i))
\end{aligned}$$

Let us prove that $g(|a-b|) = h(a,b)$, $g(0) = 0$ and the monotony of g. Let a, b, $c \in [0,1]$, and $A(x) = a$, $B(x) = b$, $C(x) = c$, $\forall x \in \Omega$.

1. $h(a,a) = 0$, $\forall a \in [0,1]$. If there exists such $a \in [0,1] \mid h(a,a) \neq 0$, then $D(A,A) = n \cdot h(a,a) \neq 0$. Absurd.
2. $h(a,b) = h(b,a)$. If there exist a, b such that $h(a,b) \neq h(b,a)$, thus $D(A,B) = n \cdot h(a,b) \neq n \cdot h(b,a) = D(B,A)$. Absurd.

3. $h(0,\cdot)$ is monotonous no decreasing in $[0,1]$. Let be $0 \leq a \leq b \leq 1$, then $\emptyset \subseteq A \subseteq B$, and $n \cdot h(0,a) = D(\emptyset, A) \leq D(\emptyset, B) = n \cdot h(0,b)$.

4. $h(a+t,b+t) = h(a,b)$, with $t \in [-1,1]$, $a+t, b+t \in [0,1]$.

   (a) if $t \leq 0$ and $a+t, b+t \in [0,1]$, let be $c = 1+t$. $D(A \cap C, B \cap C) = n \cdot h(a+c-1, b+c-1) \leq D(A,B) = n \cdot h(a,b)$. Then $h(a+t, b+t) \leq h(a,b)$.

   (b) if $t > 0$ and $a+t, b+t \in [0,1]$. Let be $c = t$, $D(A \cup C, B \cup C) = n \cdot h(a+c, b+c) \leq D(A,B) = n \cdot h(a,b)$. Then $h(a+t, b+t) \leq h(a,b)$.

   Since a, b and t are arbitrary, then it follows that $h(a+t, b+t) = h(a,b)$

5. $h(a,b) = h(|a-b|, 0)$. If $a \leq b$, then $h(a,b) = h(a-a, b-a) = h(0, b-a)$. If $b \leq a$, then $h(a,b) = h(a-b, b-b) = h(a-b, 0) = h(0, a-b)$.

6. Let be $g(a) = h(a,0)$. Then $g(0) = 0$ and $g(\cdot)$ is monotonous no decreasing in $[0,1]$.

Thus

$$D(A,B) = \sum_{i=1}^{n} h(A(x_i), B(x_i)) =$$

$$= \sum_{i=1}^{n} h(|A(x_i) - B(x_i)|, 0) =$$

$$= \sum_{i=1}^{n} g(|A(x_i) - B(x_i)|)$$

To prove the counterpart, let be $D(A,B) = \sum_{i=1}^{n} g(|A(x_i) - B(x_i)|)$, with $g(0) = 0$ and g monotonous no decreasing. Trivially, $D(A,A) = 0$ and $D(A,B) = D(B,A)$. Let be $\Omega_I = \{x \in \Omega | A(x) + C(x) \geq 1, B(x) + C(x) \geq 1\}$, $\Omega_{II} = \{x \in \Omega | A(x) + C(x) \geq 1, B(x) + C(x) < 1\}$, $\Omega_{III} = \{x \in \Omega | A(x) + C(x) < 1, B(x) + C(x) \geq 1\}$, $\Omega_{IV} = \{x \in \Omega | A(x) + C(x) < 1, B(x) + C(x) < 1\}$. Thus

$$D(A \cup C, B \cup C) =$$

$$= \sum_{x \in \Omega_I} g(|1-1|) +$$

$$+ \sum_{x \in \Omega_{II}} g(|1 - (B(x) + C(x)|) +$$

$$+ \sum_{x \in \Omega_{III}} g(|(A(x_i) + C(x)) - 1|) +$$

$$+ \sum_{x \in \Omega_{IV}} g(|(A(x) + C(x)) - (B(x) + C(x))|)$$

Since g is monotonous, then if $a + c \geq 1 \geq b + c$, it follows $g(|1 - (b+c)|) \leq g(|(a+c) - (b+c)|)$, and so on. Therefore

$$D(A \cup C, B \cup C) \leq D(A,B).$$

Similar proof for $D(A \cap C, B \cap C) \leq D(A,B)$. ■

We extend the results of [10].

Theorem 2 *Let μ be a fuzzy measure. Then there exits a divergence measure D such that*

$$\mu(A) = D(A, \emptyset)$$

Proof. Let be $(|A - B|)(x) = |A(x) - B(x)|$, $\forall x \in \Omega$, and $D(A,B) = \mu(|A - B|)$. D must verify the axioms **d1-d3** to be a divergence measure. Obviously, $D(A,B) = D(B,A)$ and $D(A,A) = 0$. To prove that $D(A \cup C, B \cup C) \leq D(A,B)$, let us see that $(|A \cup C - B \cup C|)(x) \leq (|A - B|)(x)$, $\forall x \in \Omega$. Since $(|A \cup C - B \cup C|)(x) = |\min\{A(x) + C(x), 1\} - \min\{B(x) + C(x), 1\}|$, by same reasons that in the counterpart of Theorem 1, it holds that $(|A \cup C - B \cup C|)(x) \leq (|A - B|)(x)$, $\forall x \in \Omega$. Analogous proof for $D(A \cap C, B \cap C) \leq D(A,B)$. Therefore the axiom **d3** holds and D is a divergence measure such that $\mu(A) = D(A, \emptyset)$. ■

Theorem 3 *Let D be a local divergence measure and g its generator function. Then $g(a_1 + a_2) \leq g(a_1) + g(a_2)$, $\forall a_1, a_2 \in [0,1]$, if and only if D is a pseudo-metric.*

Proof. Let us prove that

$$D(A,B) \leq D(A,C) + D(B,C).$$

Since D is local, then it must verify that

$$\sum_{i=1}^{n} g(|A(x_i) - B(x_i)|) \leq$$

$$\leq \sum_{i=1}^{n} g(|A(x_i) - C(x_i)|) + g(|C(x_i) - B(x_i)|)$$

If $C(x) \leq \min\{A(x), B(x)\}$ or $\max\{A(x), B(x)\} \leq C(x)$ it holds trivially by the monotonicity of g. If $A(x) \leq C(x) \leq B(x)$, let be $a_1 = |A(x) - C(x)|$, $a_2 = |B(x) - C(x|$. Then

$$g(|A(x) - B(x)|) \leq$$

$$\leq g(|A(x) - C(x)|) + g(|C(x) - B(x)|)$$

Analogous proof for $B(x) \leq C(x) \leq A(x)$.

The counter part is obvious ($A(x) = a_1 + a_2$, $B(x) = 0$, $C(x) = a_2$). ■

3 CONCLUSIONS

In this paper we have introduced an axiomatic approach to the cluster validity problem using basic concepts (axioms **d1-d3**) and the connectives T_∞ and S_∞.

We gave a characterization of divergence measures when all data are equally important, also we have remarked its relation with the pseudodistance and we have obtained another point of view of fuzzy measures based on divergence measures. Our interpretation says that a fuzzy measure $\mu(A)$, the degree of evidence or belief that a particular element belongs in the set A, is the divergence measure between A and $\emptyset$, the degree of evidence or belief that a particular element belongs in the set A and not in $\emptyset$.

We think that a divergence measure is suitable to quantify the difference between two fuzzy sets in the context of fuzzy partitions.

Acknowledgements

The authors are grateful to Dr. S. Montes for her help.

References

[1] Benvenutti, P., Vivona, D. & Divari, M. (1998) *Fuzziness Measures via Pseudo-Metrics on the Fuzzy Sets Space.* IPMU'98, pp.1036-1040. Paris.

[2] Burbea, B. & Rao,C.R. (1982) *Entropy differential metric, distance and divergence measures in probabilistic spaces: A unified approach.* T. Multiv. Anal., **12**, 575-596.

[3] De Baets, B. & Mesiar, R. (1998) *T-partitions.* Fuzzy Sets and Systems **97**. 213-223.

[4] Dumitrescu, D. & Pop, H.F. (1993) *Degenerate and non-degenerate convex decomposition of finite fuzzy partitions-I.* Fuzzy Sets and Systems **73**. 365-376.

[5] Dumitrescu, D. (1992) *Fuzzy partitions with the connectives T_∞, S_∞.* Fuzzy Sets and Systems **47**. 193-195.

[6] Kapur, J.N. (1994) *Measures of informations and their applications.* John Willey and Sons.

[7] Montes, S. (1998) *Particiones y medidas de divergencia en modelos difusos.* Ph. Thesis. University of Oviedo.

[8] Montes, S., Gil, P. & Bertoluzza, C.(1998) *Divergence measure between Fuzzy sets and fuzziness.* IPMU'98, Paris, 1030-1035.

[9] Pardo, L. (1997) *Teoría de la Información Estadística.* Ed. Hespérides. 1997

[10] Torres, E., Miranda, P., Gil, P. (1999) *Relations among Divergence Measures and Fuzzy Measures.* EUROFUSE-SIC'99, pp. 359-364. Budapest.

RELATIONSHIPS BETWEEN UNCERTAINTY MEASURES, FUZZINESS MEASURES AND DIVERGENCE MEASURES

Susana Montes
Dept. of Statistics and O.R.
University of Oviedo
smr@pinon.ccu.uniovi.es

Inés Couso
Dept. of Statistics and O.R.
University of Oviedo
couso@pinon.ccu.uniovi.es

Pedro Gil
Dept. of Statistics and O.R.
University of Oviedo
pedro@pinon.ccu.uniovi.es

Summary

The aim of this communication is to establish some relationships between uncertainty measures or entropies, fuzziness measures and divergence measures, that is, between different ways to measure the uncertainty, which can be produced by a random experiment and/or the inaccuracy in the definition of the subsets in the experiment. Different relationships will be studied until we establish a one-to-one function between some important families of uncertainty, fuzziness and divergence measures.

Keywords: uncertainty measure, fuzziness measure and divergence measure.

1 INTRODUCTION

Our work regards the study of uncertainty associated to systems in a fuzzy environment. The starting point of our research has been the axiomatic Information Theory of B.Forte and J.Kampé de Fèriet, where uncertainty is directly associated with a collection of (crisp) subsets of a space Ω (see for instance [1]). In the frame of this theory it is possible to guess that there exists a fairly strong relationship between uncertainty and fuzziness, and between uncertainty and classical divergence. In this respect, a fundamental work has been developed by De Luca-Termini, who introduced a kind of fuzziness measure (the non-probabilistic entropy of a fuzzy set) based on a probabilistic uncertainty measure.

In our opinion, the link between measures of uncertainty and imprecision in fuzzy environments, may lie in what we refer to as divergence measure, because of the analogy with the classical meaning of the term used by various authors ([9], ...) in comparing two probability distributions.

We will begin this work by introducing definitions and characterizations of these concepts in Section 2. Afterwards, in Section 3, we will study some relationships between uncertainty measures and fuzziness measures, and in Section 4, we will introduce these last ones and divergence measures, to end establishing an isomorphism between uncertainty, fuzziness and divergence measures.

2 PRELIMINARY

Shannon ([10]) defined the first probabilistic uncertainty measure (or entropy) in the context of Communication Theory. According to the author, the uncertainty underlying a random experiment can be measured by means of the quantity $H(P) = -\sum_{i=1}^{n} p_i log_2(p_i)$, where the values p_i are the probabilities of the possible results of that experiment. Some authors have generalized the previous measure and given other families of uncertainty measures. Salicru et al. (see [9]) proved that all these measures of entropy are part of a wider family and named them $h-\phi$–entropies. This family is slightly more general than Ben Bassat's family of f-entropies that were defined as those functions that can be expressed like $H(P) = \sum_{i=1}^{n} f(p_i)$, where f is a concave function. In another communication of us ([3]), we have introduced and characterized a family which is a bit more general than Ben Bassat's one (but different from the family of $h-\phi$–entropies) in the case of discrete distributions. We named them quasi-ϕ-entropies, and they are expressed by $H(P) = \sum_{i=1}^{n} \phi(p_i)$, where ϕ is a function such that $\phi(\lambda x+(1-\lambda)y) \geq \lambda\phi(x)+(1-\lambda)\phi(y), \forall x, y \in [0,1], x+y \leq 1$.

On the other hand, an uncertainty measure is said to verify Pigou-Dalton's condition (or Principle of Transfer) if given two probability distributions P and P' with parameters $(p_1, p_2, \ldots, p_n)$ and $(p_1', p_2', \ldots, p_n')$ respectively, then $H(P) \leq H(P')$, where, $p_k = p_k', \forall k \notin \{i, j\}$, $p_i, p_j, \delta \leq (p_i - p_j)/2, p_i' = p_i + \delta, p_j' = p_j - \delta$. In other words, the more similar the probabilities of two outcomes of the experiment are, the higher the entropy is.

Once we have established some result about uncertainty measure, we are going to introduce another very important concept, studied by many authors ([6], [11], ...), which was named fuzziness measure.

Definition 2.1
A fuzziness measure is a real function f defined on $\widetilde{P}(\Omega)$, where $\widetilde{P}(\Omega) = \{$fuzzy subsets of $\Omega\}$, such that

1. $f(\widetilde{A}) = 0 \Longleftrightarrow \widetilde{A}$ *is a crisp set.*
2. *If $\widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega)$ and $\widetilde{A}$ is "sharper" than $\widetilde{B}$, then $f(\widetilde{A}) \leq f(\widetilde{B})$.*
3. *$f(\widetilde{A})$ assumes the maximum value if and only if $\widetilde{A}$ is "maximally fuzzy".*

Depending on the fuzziness degree meaning and measurement, the terms "sharper than" and "maximally fuzzy" will be considered in different ways. We will work with these two criteria, $\widetilde{A}$ is said to be sharper than $\widetilde{B}$ if and only if

1. either $A(x) \leq B(x) \leq 1/2$ or $A(x) \geq B(x) \geq 1/2, \forall x \in \Omega$ (see [5]),
   or more generally,
2. $\widetilde{A}$ is said to be sharper than $\widetilde{B}$ if and only if $|\widetilde{A}(x) - 1/2| \leq |\widetilde{B}(x) - 1/2|, \forall x \in \Omega$ (see [4]).

A very important family of fuzziness measure is the Knopfmacher's class ([6]), that is, the family formed by the functions f such that

$$f(\widetilde{A}) = F\left(\sum_{w \in \Omega} c_w \cdot g_w(\widetilde{A}(w))\right), \forall \widetilde{A} \in \widetilde{P}(\Omega),$$

where $c_w \in \mathbb{R}^+$; g_i is a real-valued function such that $g_w(0) = g_w(1) = 0, g_w(t) = g_w(1-t), \forall t \in [0, 1]$ and g_w is strictly increasing on $[0, 1/2]$; F is a positive strictly increasing function with $F(0) = 0$.

In other works ([7], [8]) we have considered the particular Knopfmacher fuzziness measure such that F is the identity, g_w is the same for all $w \in \Omega$ (we have denoted g_w by u_f or simply u) and u is concave. Any function in this family was named local fuzziness measure.

Finally, we will introduce the last concept that we are going to relate to the two previous ones, which was studied firstly in [8].

Definition 2.2 *([7]) Let Ω be the referential. A map $D : \widetilde{P}(\Omega) \times \widetilde{P}(\Omega) \longrightarrow \mathbb{R}$ is a divergence measure if and only if $\forall \widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega)$, D satisfies the following conditions:*

1. $D(\widetilde{A}, \widetilde{B}) = D(\widetilde{B}, \widetilde{A})$;
2. $D(\widetilde{A}, \widetilde{A}) = 0$;
3. $\max\{D(\widetilde{A} \cup \widetilde{C}, \widetilde{B} \cup \widetilde{C}), D(\widetilde{A} \cap \widetilde{C}, \widetilde{B} \cap \widetilde{C})\} \leq D(\widetilde{A}, \widetilde{B}), \forall \widetilde{C} \in \widetilde{P}(\Omega)$.

According to this definition, if we consider only the case where Ω is finite, since the couples $(\widetilde{A}, \widetilde{B}), (\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\})$ only differ in the ith element, it seems natural to suppose that the variation of divergence only depends on what has been changed, and according to this, it is said that a divergence measure on a finite referential is local if and only if for all $\widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega)$ we have that $D(\widetilde{A}, \widetilde{B}) - D(\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\}) = h(\widetilde{A}(x_i), \widetilde{B}(x_i))$.

The following statement characterizes the local divergence measures.

Proposition 2.3 *A map D is a local divergence if and only if there exists a function $h_D : [0, 1] \times [0, 1] \longrightarrow \mathbb{R}$ such that*

$$D(\widetilde{A}, \widetilde{B}) = \sum_{w \in \Omega} h_D(\widetilde{A}(w), \widetilde{B}(w))$$

where

i) $h_D(x, y) = h_D(y, x), \forall x, y \in [0, 1]$;

ii) $h_D(x, x) = 0, \forall x \in [0, 1]$;

iii) $h_D(x, z) \geq \max\{h_D(x, y), h_D(y, z)\}$, $\forall x, y, z \in [0, 1]$ with $x \leq y \leq z$.

3 UNCERTAINTY MEASURES AND FUZZINESS MEASURES

From now on, we are going to consider that the referential Ω is finite.

Proposition 3.1 *Let H be an uncertainty measure that verifies the Pigou-Dalton's condition and such that*

$H(P) = 0 \iff P$ *is degenerate. Let us consider a function* $u_H : [0,1] \longrightarrow I\!R^+$ *such that* $u_H(x) = H(x, 1-x), \forall x \in [0,1]$. *Then, the map* $f : \widetilde{P}(\Omega) \longrightarrow I\!R^+$ *defined by* $f(\widetilde{A}) = h\left(\sum_{w\in\Omega} u_H(\widetilde{A}(w))\right)$, *where* $h : I\!R^+ \longrightarrow I\!R^+$ *is a non-decreasing function such that* $h(x) = 0 \iff x = 0$, *is a fuzziness measure according to the first definition of "fuzzy subset sharper than"and it belongs to Knopfmacher's class.*

Proof. It is a direct consequence of Theorem 2.1 in [2]. ■

The uncertainty measures more widely used are symmetrical, and so, if they have this property, the fuzziness measures they generate are according to the second definition of "fuzzy subset sharper than", which is stronger than the first one and we also think that is more natural.

Proposition 3.2 *In the preceding proposition, if* H *has also a symmetry property, that is, if* $H(x, 1-x) = H(1-x, x), \forall x \in [0,1]$, *then* f *is also a fuzziness measure according to the second definition "fuzzy subset sharper than".*

Proof. This is again a direct consequence of Theorem 2.1 in [2]. ■

Thus, if we have a uncertainty measure H satisfying adequate conditions, we can construct a fuzziness measure with this one.

Lemma 3.3 *A uncertainty measure* H *is a quasi-*ϕ*-entropy if and only if* H *is symmetrical, expansible, branching, continue and it fulfils Pigou-Dalton's condition.*

Proof. To prove this result, it is necessary to use Proposition 2.3 in [3]. ■

Once we have characterized the quasi-ϕ-entropies, we can use this result to establish a one-to-one function between uncertainty and fuzziness measures.

Proposition 3.4 *Let* $F_1 : \mathcal{H} \longrightarrow \mathcal{F}$ *be a function such that* $F_1(H)(\widetilde{A}) = \sum_{w\in\Omega} u_H(\widetilde{A}(w))$, *where* u_H *is the function defined in Proposition 3.1,* $\mathcal{H}$ *is a family of symmetrical and continue uncertainty measures satisfying Pigou-Dalton's condition and such that* $H(P) = 0 \iff P$ *is degenerate, and* $\mathcal{F}$ *is a family of fuzziness measures. The function* F_1 *is injective if and only if any uncertainty measure in* $\mathcal{H}$ *is a quasi-*ϕ *entropy with* ϕ *continue,* $\phi(x) = \phi(1-x), \forall x \in [0, \frac{1}{2}]$ *and* $\phi(x) = 0 \iff x = 0$.

Proof. In this proof we have to consider Proposition 3.2 to see that F_1 is actually a function. Besides, $F_1(H)$ is a local fuzziness measure and then it is easy to prove that F_1 is injective. In order to prove the converse, we need to use Lemma 3.3. ■

Theorem 3.5 *If* $\mathcal{H}$ *is the class of* ϕ*-entropies such that* ϕ *is symmetrical with respect to* $1/2$ *and* $H(P) = 0 \iff P$ *is degenerate and if* $\mathcal{F}$ *is the class of local fuzziness, then the function* F_1 *defined in Proposition 3.4 is a one-to-one function.*

Proof. If H is a ϕ-entropy, then H is a quasi-ϕ-entropy, and then we have only to prove, by using Proposition 3.4, that F_1 is surjective. In order to see this, we have to consider the ϕ-entropy H defined by $\phi(x) = 1/2u_f(x), \forall x \in [0,1]$, and then $F_1(H) = f$. ■

4 FUZZINESS MEASURES AND DIVERGENCE MEASURES

In order to relate fuzziness measures and divergence measures, to continue with the ideas of fuzziness measures proposed by Yager ([12]), we need to establish a previous lemma for non-differentiable functions.

Lemma 4.1 *If a function* $u : [0,1] \longrightarrow I\!R^+$ *is concave, then the function* $h : [0,1] \times [0,1] \longrightarrow I\!R^+$ *defined by* $h(x,y) = u\left(\frac{x+y}{2}\right) - \frac{u(x)+u(y)}{2}$ *satisfies that* $h(\cdot, y)$ *is a non-increasing function in* $[0,y]$ *and a non-decreasing function in* $[y,1]$, *for any* y *fixed in* $[0,1]$.

Proof. To prove this lemma it is necessary to consider that if u is concave in $(0,1)$, then u is continue, there exist $u'(x^+), u'(x^-), \forall x \in (0,1)$, and for all $x_1, x_2 \in (0,1)$ such that $x_1 < x_2$, then $u'(x_1^-) \geq u'(x_1^+) \geq u'(x_2^+) \geq u(x_2^-)$ and we also have to prove a result similar to Lagrange's Theorem for functions r such that there exists $r(x^+)$ and $r(x^-)$ for all x, but r is non-differentiable. ■

By considering this result and the properties of local fuzziness measure, we have that

Proposition 4.2 *If we consider the function* F_2 *defined in* $\mathcal{F}$ *(family of local fuzziness measures) by the map* $F_2(f) : \widetilde{P}(\Omega) \times \widetilde{P}(\Omega) \longrightarrow I\!R$ *with*

$$F_2(f)(\widetilde{A}, \widetilde{B}) = f\left(\frac{\widetilde{A}+\widetilde{B}}{2}\right) - \frac{f(\widetilde{A})+f(\widetilde{B})}{2},$$
$$\forall \widetilde{A}, \widetilde{B} \in \widetilde{P}(\Omega),$$

then $F_2(f)$ is a local divergence measure such that $h(x, 1-x)$ is a convex function of x and

$$h_D(x,y) = \tfrac{1}{2}h_D(x,1-x) + \tfrac{1}{2}h_D(y,1-y) - h_D\left(\tfrac{x+y}{2}, 1-\tfrac{x+y}{2}\right), \forall x,y \in [0,1].$$

Proof. The function $h_{F_2(f)}$ which define the local divergence measure is $h_{F_2(f)}(x,y) = u_f\left(\frac{x+y}{2}\right) - \frac{u(x)+u(y)}{2}, \forall x,y \in [0,1]$. By applying Lemma 4.1, we can prove that $F_2(f)$ defined by this $h_{F_2(f)}$ satisfies the condition in Proposition 2.3. It is trivial that $h_{F_2(f)}$ verifies the condition in the proposition. ■

Thus, the local divergence measure considered in the previous proposition is completely characterized if we know the value of $h(x, 1-x), \forall x \in [0,1]$.

Theorem 4.3 *If $\mathcal{D}$ is the class of local divergence measures such that $h_D(x,y) = \frac{1}{2}h_D(x,1-x) + \frac{1}{2}h_D(y,1-y) - h_D\left(\frac{x+y}{2}, 1-\frac{x+y}{2}\right), \forall x,y \in [0,1]$ and $h(x,1-x)$ is a convex function of x, then the map $F_2 : \mathcal{F} \longrightarrow \mathcal{D}$ is a one-to-one function.*

Proof. F_2 is a function by applying Proposition 4.2. It is injective since $h_{F_2(f)}$ is completely characterized by the values of $h_{F_2(f)}(x,1-x) = u_f(1/2) - u_f(x), \forall x \in [0,1]$. F_2 is surjective since for all $D \in \mathcal{D}$ we can consider de local fuzziness measure defined by $u_f(x) = h_D(1,0) - h_D(x,1-x), \forall x \in [0,1]$, and then $f_2(f) = D$. ■

By joining Theorems 3.5 and 4.3 we have that

Corollary 4.4 *The set $\mathcal{H}$ of the class of quasi-ϕ entropy with ϕ continue, $\phi(x) = \phi(1-x), \forall x \in [0,1]$ and $\phi(x) = 0 \iff x = 0$, the set $\mathcal{F}$ of local fuzziness measures and the set $\mathcal{D}$ of the local divergence measures such that $h(x,y) = \frac{1}{2}h(x,1-x) + \frac{1}{2}h(y,1-y) - h\left(\frac{x+y}{2}, 1-\frac{x+y}{2}\right), \forall x,y \in [0,1]$ and $h(x,1-x)$ is a convex function of x, are isomorphic.*

5 CONCLUSIONS AND OPEN PROBLEMS

In this paper we have studied some relations between different ways to measure the uncertainty (probabilistic or non-probabilistic) and a measure of the difference between two fuzzy subsets. The main result is Corollary 4.4, where we have given a way to construct a divergence measure if we have an uncertainty measure or a fuzziness measure, and conversely.

With regard to the open problems, the most immediate is to find other relationships between this functions and other different ways to measure the uncertainty (fuzzy measure, aggregation operators, etc).

Acknowledgements

The research in this paper is supported in part by DGICYT Grant No. PB97-1286. Its financial support is gratefully acknowledged.

References

[1] C. Bertoluzza and M. Schneider (1974). Informations totalement composables. *Théories de l'Information, Lect. Notes Math.* **398**, pp 90–98.

[2] I. Couso, S. Montes and P. Gil (1996). Measures of fuzziness and Information Theory. *Proceedings of IPMU'96 Conference* pp 1030–1035.

[3] I. Couso and P. Gil (1998). Characterization of a family of entropy measures. *Proceedings of IPMU'98 Conference* pp 1053–1059.

[4] D. Dubois and H. Prade (1980). Fuzzy Sets and Systems. Theory and Applications. *Academic Press, Inc..*

[5] G.J. Klir and T.A. Folger (1988). Fuzzy Sets, Uncertainty, and Information. *Prentice-Hall International (UK) Limited.*

[6] J. Knopfmacher (1975). On measures of fuzziness. *J. Math. Analysis and Applications* **49**, pp529–534.

[7] S. Montes, P. Gil and C. Bertoluzza. Divergence between fuzzy sets and fuzziness (1998). *Proceedings of IPMU'98 Conference* pp 1030–1035.

[8] S. Montes and P. Gil (1998). Some Classes of Divergence Measures Between Fuzzy Subsets and Between Fuzzy Partitions. *Mathware and Soft Computing* **5**, pp 253–265.

[9] L. Pardo (1997). Teoría de la Información estadística. Ed. Hesperides, Salamanca.

[10] C.E. Shannon (1948). A mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, pp 379–423, 623–656.

[11] E. Trillas and T. Riera (1978). Entropies in Finite Fuzzy Sets. *Inf. Sc.* **15**, 159–168.

[12] R.R. Yager (1995). Measures of Entropy and Fuzziness related to Agregation Operators. *Inf. Sc.* **82**, 147–166.

LOCAL DIVERGENCE MEASURES ON INFINITE REFERENTIALS

Jorge Jiménez
Dept. of Mathematics
University of Oviedo
meana@etsiig.uniovi.es

Susana Montes
Dept. of Statistics and O.R.
University of Oviedo
smr@pinon.ccu.uniovi.es

Carlo Bertoluzza
Dept. of Computer Eng. and Systems
University of Pavia
carlo@unipv.it

Summary

The aim of this work is to define and to characterize the local divergence measures when working with infinite referentials , that is, those measures which allow us to study the degree of similarity between two fuzzy sets which are decomposables. Once this concept is characterized with several different definitions, some of their properties are studied and some examples are presented.

Keywords: measurable space, measure σ-additive, divergence measure.

1 INTRODUCTION

In several previous papers ([6], [7], ...), it emerged a new form to quantify the degree of difference between two fuzzy subsets within any referential Ω from a function, which will be defined below, and it will be called divergence measure, which has as particular cases the customary distances between fuzzy sets already known. This new concept emerged in a sense as an extension of the classic probabilistic divergence measures ([8], ...), taking advantage of the strong relationship existing between some concepts of the Information Theory and the Fuzzy sets Theory as, for example, the probabilistic entropy of Shannon ([9]) and the non-probabilistic entropy of De Luca & Termini ([2]).

Once the definition of divergence measure was given, special emphasis was made on the study of the defined divergences on finite spaces which could be decomposed as the sum of the value of a function in each point of the referential . The extension of all this theory to numerable referentials was a simple exercise, but not thus the extension to no numerable infinite spaces, which will go through the measure spaces' study, and where the proofs, though they could seem a simple generalization, need an analytical rigor that will make them considerably more complex. This is precisely the study made in this paper, in which we find a Section 2 that presents the measure definition of local divergence for a non-numerable infinite referential and are studied several related results; in Section 3 we study some interesting properties of the measures defined in the previous section, these properties, as well as all the results of the Section 2, do not include the proofs which are sometimes very long, but we simply make some short commentaries on the most relevant parts. Finally, in Section 4, possible applications and future projects related to this one are commented.

2 LOCAL DIVERGENCE MEASURES

The measure of the difference of two fuzzy subsets is defined axiomatically on the basis of the following natural properties:

- It is a non-negative and symmetric function of the two fuzzy subsets to be compared.
- It becomes zero when the two sets coincide.
- It decreases as the two subsets become "more similar" in some sense.

Whereas it is easy to formulate analytically the first and the second condition, the third one depends on the formalization of the concept of "more similar". We base our approach on the fact that if we add (in the sense of union) a subset $\widetilde{C}$ to both fuzzy subsets $\widetilde{A}$ and $\widetilde{B}$, we obtain two subsets which are closer to each other; the same happens with the intersection.

Before giving the divergence measure definition, we need to include a previous lemma, for this definition to make sense.

Lemma 2.1 *Let $(\Omega, \mathcal{A})$ be a measure space, where Ω is any set and $\mathcal{A}$ is a σ-field of crisp sets of Ω, and let $\widetilde{P}(\Omega)$ be the set of all the fuzzy subsets of Ω, then the*

set $\mathcal{A}^*$ defined by

$$\mathcal{A}^* = \{\widetilde{A} \in \widetilde{P}(\Omega) / \text{ the membership function of } \widetilde{A}, \widetilde{A} : (\Omega, \mathcal{A}) \longrightarrow ([0,1], \beta) \text{ is a measurable function}\}$$

is a σ-field over Ω.

Proof. The proof is a consequence of the fact that the Borel's σ-field (β) is generated by the real intervals, $\mathcal{A}$ is a σ-field and that the definitions of union and intersection which we use, are the maximum and the minimum considered by Zadeh. ■

So we propose the following

Definition 2.2 *Let Ω the universe under study, and let $\mathcal{A}$ be a σ-field defined on Ω (we can consider in particular the family of the fuzzy subset of Ω, that is, $P(\Omega)$). A map $D : \mathcal{A}^* \times \mathcal{A}^* \longrightarrow \mathbb{R}$ is a divergence measure if and only if $\forall \widetilde{A}, \widetilde{B} \in \mathcal{A}^*$, satisfies the following conditions:*

1. $D(\widetilde{A}, \widetilde{B}) = D(\widetilde{B}, \widetilde{A})$;
2. $D(\widetilde{A}, \widetilde{A}) = 0$;
3. $\max\{D(\widetilde{A} \cup \widetilde{C}, \widetilde{B} \cup \widetilde{C}), D(\widetilde{A} \cap \widetilde{C}, \widetilde{B} \cap \widetilde{C})\} \leq D(\widetilde{A}, \widetilde{B}), \forall \widetilde{C} \in \mathcal{A}^*$.

According to this definition, if we consider only the case where Ω is finite, since the couples $(\widetilde{A}, \widetilde{B}), (\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\})$ only differ in the ith element, it seems natural to suppose that the variation of divergence only depends on what has been changed, and according to this, it is said that a divergence measure on a finite space is local, if and only if, $\forall \widetilde{A}, \widetilde{B} \in P(\Omega)$ and $\forall x_i \in \Omega$ we have that $D(\widetilde{A}, \widetilde{B}) - D(\widetilde{A} \cup \{x_i\}, \widetilde{B} \cup \{x_i\}) = h(\widetilde{A}(x_i), \widetilde{B}(x_i))$. To extend this definition to the general case we need the following result:

Lemma 2.3 *Let $(\Omega, \mathcal{A})$ be a measurable space, and $P(\Omega)$ the set formed by all the crisp subsets of Ω, then we have that $P(\Omega) \cap \mathcal{A}^* = \mathcal{A}$, that is, all the elements of the σ-field belong to $\mathcal{A}^*$, all the crisp elements of $\mathcal{A}^*$ belong to $\mathcal{A}$.*

Proof. The proof is followed from the fact that, $\mathcal{A}$ is a σ-field and that the inverse image of the intervals $(-\infty, z]$ by the membership function of a crisp subset, A is either the empty set, or Ω, or A^C. ■

From now on, we will consider the measure space $(\Omega, \mathcal{A}, \mu)$ ([3]), where Ω is a infinite referential such that $\mu(\Omega) < \infty$.The extension of the local divergence measure to this kind of referentials can't be made point by point, so we have to extend this concept subset by subset.Thus,

Definition 2.4 *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. A divergence measure D, is said to be local if and only if there exists a function $h : [0,1] \times [0,1] \longrightarrow \mathbb{R}$ measurable respect to the Borel's σ-fields such that*

$$D(\widetilde{A}, \widetilde{B}) - D(\widetilde{A} \cup Z, \widetilde{B} \cup Z) = \int_Z h(\widetilde{A}(w), \widetilde{B}(w)) d\mu(w),$$

for all $Z \in \mathcal{A}$ and for any $\widetilde{A}$ y $\widetilde{B} \in \mathcal{A}^*$, where $\int$ denote the Lebesgue-Stielges integral.

All the integrals which we calculate in the previous definition are well defined by applying the lemmas 2.1 and 2.3.

Before going to the next result, we must see a previous lemma which will be necessary for its proof.

Lemma 2.5 *Let $\widetilde{A}$ be a fuzzy subset of Ω such that $\widetilde{A}(w)$ is constant for all $w \in \Omega$, then $\widetilde{A}$ is measurable with regard to the σ-field $\mathcal{A}^*$.*

Proof. The proof is followed from the fact that the inverse image of the intervals $(-\infty, z]$ by the membership function of a crisp subset, A is either the empty set, or Ω. ■

The following statement characterizes the local divergence measures.

Proposition 2.6 *Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space. If a map $D : \mathcal{A}^* \times \mathcal{A}^* \longrightarrow \mathbb{R}$ is a local divergence then there exists a function $h : [0,1] \times [0,1] \longrightarrow \mathbb{R}$ such that*

$$D(\widetilde{A}, \widetilde{B}) = \int_\Omega h(\widetilde{A}(w), \widetilde{B}(w)) d\mu(w)$$

where
i) $h(x,y) = h(y,x), \forall x, y \in [0,1]$;
ii) $h(x,x) = 0, \forall x \in [0,1]$;
iii) $h(x,z) \geq \max\{h(x,y), h(y,z)\}, \forall x, y, z \in [0,1]$ with $x \leq y \leq z$.
iv) h is measurable.
Conversely, If D is a function defined by $D(\widetilde{A}, \widetilde{B}) = \int_\Omega h(\widetilde{A}(w), \widetilde{B}(w)) d\mu(w)$ where h verify the four before properties then h is a local divergence measure.

Proof. To prove the fist implication is necessary to consider in Definition 2.4 $Z = \Omega$, and apply Lemma 2.5 together with the fact that $\mu(\Omega) < \infty$. we also have to define the constant sets $\widetilde{A}(w) = x, \widetilde{B}(w) = y$ y $\widetilde{C}(w) = z$ for all $w \in \Omega$ and to consider that the Lebesgue-Stielges integral is only defined for measurable functions, with the aim to prove the four properties of h.

In the converse proof we have to consider the fact that the Lebesgue-Stielges integral of a positive measurable function always exists, the properties of h, and the decomposition of Ω in the five following disjoined subsets: $\Omega_1 = \{w \in \Omega / \widetilde{A}(w) \leq \widetilde{B}(w) \leq \widetilde{C}(w)\ \ \widetilde{B}(w) < \widetilde{A}(w) \leq \widetilde{C}(w)\}$, $\Omega_2 = \{w \in \Omega / \widetilde{A}(w) \leq \widetilde{C}(w) < \widetilde{B}(w)\}$, $\Omega_3 = \{w \in \Omega / \widetilde{B}(w) \leq \widetilde{C}(w) < \widetilde{A}(w)\}$, $\Omega_4 = \{w \in \Omega / \widetilde{C}(w) < \widetilde{A}(w) \leq \widetilde{B}(w)\}$ y $\Omega_5 = \{w \in \Omega / \widetilde{C}(w) < \widetilde{B}(w) < \widetilde{A}(w)\}$, which are measurable subsets, because of being measurable a continuous function of measurable functions. ■

We can use the previous proposition to generate divergence measures easily in infinite referentials. The importance of this fact leads us to the searching of others possible conditions for the function h, which will be equivalent. So we have the following result

Corollary 2.7 *Condition iii) in Proposition 2.6 for h can be replaced by $h(\cdot, y)$ is a function decreasing in $[0, y]$ and increasing in $[y, 1]$.*

Up to now, We have seen, a local divergence measure based on the union with a crisp set. In the following result, we are going to give an equivalent condition based on the intersection with a crisp set.

Lemma 2.8 *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, if D is a local divergence measure, then for all crisp subset Z of Ω it is verified that*

$$D(\widetilde{A} \cup Z, \widetilde{B} \cup Z) = D(\widetilde{A} \cap Z^c, \widetilde{B} \cap Z^c), \forall \widetilde{A}, \widetilde{B} \in \mathcal{A}^*.$$

Proof. To prove this lemma is necessary to apply the definition of local divergence measure to the sets $\widetilde{A} \cap Z^c$ and $\widetilde{B} \cap Z^c$ with respect to Z. ■

Proposition 2.9 *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, if D is a divergence measure, then D is local if and only if there exists a measurable function $h : [0,1] \times [0,1] \longrightarrow \mathbb{R}$ such that*

$$D(\widetilde{A}, \widetilde{B}) - D(\widetilde{A} \cap Z, \widetilde{B} \cap Z) = \int_{Z^c} h(\widetilde{A}, \widetilde{B}) d\mu,$$

$\forall \widetilde{A}, \widetilde{B} \in \mathcal{A}^*$ and $\forall Z \in P(\Omega)$.

Proof. In the proof we have to use the previous lemma and the definition of local divergence measure applied to Z^c. ■

To end this section, we will enunciate a result which will allow us to generate divergence measures from a local divergence measure.

Proposition 2.10 *Let D be a local divergence measure associated to a measurable function h, if $\phi : \mathbb{R} \longrightarrow \mathbb{R}$ is a non-decreasing function and $\phi(0) = 0$, then the maps D_ϕ and D^ϕ are divergence measures, and, besides, D_ϕ is local, where*

$$D_\phi(\widetilde{A}, \widetilde{B}) = \int_\Omega \phi(h(\widetilde{A}(w), \widetilde{B}(w)) d\mu(w),$$
$$D^\phi(\widetilde{A}, \widetilde{B}) = \phi\left(\int_\Omega h(\widetilde{A}(w), \widetilde{B}(w) d\mu(w)\right).$$

Proof. To prove this proposition, we must take into account that $(\phi \circ h)^{-1}$ is equal to either$((-\infty, z]) = h^{-1}((-\infty, k_1])$ or $h^{-1}((-\infty, k_1))$ or $h^{-1}(\mathbb{R})$, where $k_1 = \sup\{k \in \mathbb{R} / \phi(k) \leq z\}$, and that any of these three sets belong to the σ-field $\mathcal{A}$. ■

3 PROPERTIES OF THE LOCAL DIVERGENCE MEASURES

We will study several properties which are verified by the local divergence measures. This properties relate them to the fuzziness degree of the sets. Before stating the different properties we must explain the notation. $\widetilde{A} \lll \widetilde{B}$ will denote that couple of sets such that verify $|\widetilde{A}(w) - 1/2| \geq |\widetilde{B}(w) - 1/2|, \forall w \in \Omega$, and $\widetilde{A} \ll \widetilde{B}$ will denote that couple of sets such that verify $\widetilde{A}(w) \leq \widetilde{B}(w) \leq 1/2$ o $\widetilde{A}(w) \geq \widetilde{B}(w) \geq 1/2, \forall w \in \Omega$.

Proposition 3.1 *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space, and let $\widetilde{A}$, $\widetilde{B}$ be in $\mathcal{A}^*$. Given D, a divergence measure with the local property, we have that*

1. *If $\widetilde{A} \lll \widetilde{B}$ then $D(\widetilde{A}, \widetilde{A}^c) \geq D(\widetilde{B}, \widetilde{B}^c)$.*
2. *$D(Z, Z^c) = D(V, V^c), \forall Z, V \in P(\Omega)$.*
3. *$D(\widetilde{A}, \widetilde{A}) \leq D(\widetilde{A}, \widetilde{B}) \leq D(Z, Z^c), \forall Z \in P(\Omega)$.*
4. *If the function h which generates D satisfies that $h(x, e) = h(c(x), e), \forall x \in [0,1]$, where e is the fixed point of complementary function ^{c}, then $D(\widetilde{A}, \widetilde{E}_c) = D(\widetilde{A}^c, \widetilde{E}_c)$, where $\widetilde{E}_c(w) = e, \forall w \in \Omega$.*
5. *Given $\widetilde{A}, \widetilde{B}, \widetilde{C} \in \mathcal{A}^*$ such that for all $w \in \Omega$ we have that either $\widetilde{A}(w) \leq \widetilde{B}(w) \leq \widetilde{C}(w)$ or $\widetilde{A}(w) \geq \widetilde{B}(w) \geq \widetilde{C}(w)$, then it is got that $D(\widetilde{A}, \widetilde{C}) \geq \max\{D(\widetilde{A}, \widetilde{B}), D(\widetilde{B}, \widetilde{C})\}$.*
6. *If $\widetilde{A} \ll \widetilde{B}$ then $D(\widetilde{A}, \widetilde{E}) \geq D(\widetilde{B}, \widetilde{E})$ where $\widetilde{E}$ is the fuzzy set defined in property 4 when $c(x) = 1 - x$.*
7. *The function h which generates D, satisfies that $h(\cdot, 1/2)$ is symmetrical with regard to $1/2$ if and only if when $\widetilde{A} \lll \widetilde{B}$ then $D(\widetilde{A}, \widetilde{E}) \geq D(\widetilde{B}, \widetilde{E})$.*

8. *The function h which generates D satisfies that $h(x,0) = h(1-x,1), \forall x \in [0,\frac{1}{2}]$, if and only if $\widetilde{A} << \widetilde{B}$ implies that $D(\widetilde{A}, N_{\widetilde{A}}) \leq D(\widetilde{B}, N_{\widetilde{B}})$, where $N_{\widetilde{A}}, N_{\widetilde{B}}$ represent the nearest crisp set to $\widetilde{A}$ and $\widetilde{B}$, respectively.*

9. *If $h(\cdot, 1/2)$ is symmetric with regard to $1/2$, then $D(Z,\widetilde{E}) = D(V,\widetilde{E}), \forall Z, V \in P(\Omega)$.*

10. *If $\widetilde{A} \ll\ll \widetilde{B}$, we have that $D(\widetilde{A},\widetilde{E}) \geq D(\widetilde{B},\widetilde{E})$, if and only if the function h associated with D satisfies that $h(\cdot, 1/2)$ is symmetric with regard to $1/2$.*

Proof. To prove the first property is necessary to decompose Ω into three disjoined sets in the adequate way. These sets will be measurable by applying again the following property "a continuous function composed with measurable functions is measurable.
In the second property, we get that the divergence between a crisp set and its complementary is always constant and equal to $h(1,0) \cdot \mu(\Omega)$.
The third and fourth property are trivial by using the monotony of h and the hypothesis of the respective statements.
In the fifth one it is necessary to decompose again Ω in three measurable disjoined sets and to use the monotony of h.
The sixth one is consequence of the previous one.
In the seventh one, we have to consider the set $\widetilde{A}$ defined by $\widetilde{A}(w) = x, \forall w \in \Omega$ which verifies that $\widetilde{A} \ll\ll \widetilde{A}^c$ and $\widetilde{A}^c \ll\ll \widetilde{A}$, and we have to decompose Ω again in the four adequate measurable and disjoined sets.
The eighth one is proved using the previous decomposition of Ω and the monotony of the integral.
The ninth one is a particular case of the tenth one which we can prove using the fuzzy set $\widetilde{A}(w) = x\ \forall x \in \Omega$ and its complementary, the monotony and symmetry of h. ■

4 CONCLUSIONS AND OPEN PROBLEMS

In this paper we have introduced and studied the concept of local divergence measures defined on a infinite referentials. Its importance lies on the fact that all the applications of the local divergence measures can now be used in the infinite case, which presents some existence and integrability problems which are erased when defining this kind of measures in a coherent way.

We have the following open problems: First, to study the fuzziness measures on infinite referentials which generate these divergence measures. This first problem, has already been analysed and we obtain very satisfactory results which allow us to generate some of the already known fuzziness measures in the infinite case (Minkowski ([5]), Knopfmacher ([4]), etc), and this will be the target of a future paper.

Other immediate open problems will be the extension of all the studies made in the finite referential case to the infinite case.

As further purposes we have the composition of a parallel theory in the finite and infinite case of the studies related to the divergence measures which we are making.

Acknowledgements

The research in this paper is supported in part by DGICYT Grant No. PB97-1286. Its financial support is gratefully acknowledged.

References

[1] P. Billingsley (1979). Probability and Measure. *John Wiley & Sons, Inc.*.

[2] A. De Luca, and S. Termini (1972). A definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory. *Inf. and Control* **20**, pp 301-312.

[3] M. Loevè. Probability theory (1963). *Van Nostrand Reinhold company.*

[4] D. Dubois and H. Prade (1980). Fuzzy Sets and Systems. Theory and Applications. *Academic Press, Inc.*.

[5] G.J. Klir and T.A. Folger (1988). Fuzzy Sets, Uncertainty, and Information. *Prentice-Hall International (UK) Limited.*

[6] S. Montes and P. Gil (1998). Some Classes of Divergence Measures Between Fuzzy Subsets and Between Fuzzy Partitions. *Mathware and Soft Computing* **5**, pp 253-265.

[7] S. Montes, P. Gil and C. Bertoluzza (1998). Divergence between fuzzy sets and fuzziness. *Proceedings of IPMU'98 Conference*, pp 1030-1035.

[8] L. Pardo. Teoría de la Información estadística (1997). Ed. Hesperides, Salamanca.

[9] C.E. Shannon (1948). A mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, pp 379-423, 623-656.

On measuring μ-T-conditionality of fuzzy relations (Part II)

Luis Garmendia
DB/DC
IBM Learning Services
E-mail:
83889573@es.ibm.com

Enric Trillas
Dep. Inteligencia Artificial
Facultad de Informática
Univ. Politécnica de Madrid
E-mail: etrillas@fi.upm.es

Adela Salvador
Dep. Matemática Aplicada
E.T.S.I. Caminos
Univ. Politécnica de Madrid
E-mail: ma08@dumbo.caminos.upm.es

Abstract

This paper continues and finishes other works toward proposing a method to measure the μ-T-conditionality character of any fuzzy relation for any continuous t-norm T.

Keywords: fuzzy relation, conditionality, conditionalized fuzzy relation, conditionality measure of a fuzzy relation.

1 INTRODUCTION

The μ-T-conditionality property of fuzzy relation seems to be very interesting in order to make fuzzy inference that generalises the Modus Ponens property.

Let E_1, E_2 be two sets, let E be the set $E_1 \cup E_2$, and let μ: $E \to [0,1]$ be a fuzzy set and let T be a continuous t-norm. A fuzzy relation R: $E_1 \times E_2 \to [0,1]$ holds the μ-T-conditionality property if and only if $T(\mu(a), R(a,b)) \leq \mu(b)$ for all (a, b) in $E_1 \times E_2$.

There were proposed [1] two ways for measuring μ-T-conditionality of fuzzy relations. A first way calculates a generalised distance between a fuzzy relation R and the greatest μ-T-conditional relation that is contained in R. The other way measures the difference between $T(\mu(a), R(a,b))$ and $\mu(b)$ in all (a,b) in which R is not μ-T-conditional.

This paper completes a partial result obtained in [2] which proves that when T is the t-norm minimum or any archimedean t-norm, and when a generalised distance defined from a residuated operator of the T-norm is used, both methods proposed in [1] give the same measure values of μ-T-conditionality of fuzzy relations. This paper proves such result also for ordinal sums, so the result is finally held for all continuous t-norms.

2 PRELIMINARIES

1. Let T_R^{μ} be the fuzzy relation defined by $T_R^{\mu}(a,b)=T(\mu(a), R(a,b))$.

2. From the well known operation in [0,1], $J^T(x, y)=\mathrm{Sup}\{z / T(x, z) \leq y\}$, lets define J_μ^T as the fuzzy relation $J_\mu^T(a, b)=J^T(\mu(a), \mu(b))$.

3. The **μ-T-conditionalized relation** of R is defined as:

$$\mathbf{R_C^{\mu\text{-}T}(a,b) = Min(R(a,b), J_\mu^T(a,b))}$$

$$= \begin{cases} R(a, b) & \text{if } T_R^{\mu}(a, b) \leq \mu(b) \\ J_\mu^T(a, b) & \text{otherwise} \end{cases}$$

4. A Generalised Metric Space is a triplet (E, $\wp$, m) where E is a set, $\wp$=(A, S, ≤, e) is a commutative semigroup with neutral element e, and m is a **S-distance**, which is an application m: E×E→A such that:

1) m(a, a) = e, for all a in E

2) m(a, c) ≤ S(m(a, b), m(b, c)), for all a,b,c in E (S-triangular inequality)

Let T* be the dual t-conorm of a t-norm T, defined by T*(x,y)=1-T(1-x, 1-y).

Let $\Im^*$=([0, 1], T*, ≤, 0) be a commutative ordered semigroup with neutral element 0.

Given a relational structure (E, J), if J is a T-preorder then the function d(a, b)=1- $J^T(a, b)$ is a T*-distance in the generalised metric space ([0,1], $\Im^*$, d).

5. It is defined the μ-T-unconditionality region of a fuzzy relation, $\mathbf{INC_T^{\mu}(R)}$, as the subset of $E_1 \times E_2$ in which R is not μ-T-conditional, that is:

$\mathbf{INC_T^{\mu}(R)}$ = {(a, b)∈ $E_1 \times E_2$ such that $T(\mu(a), R(a, b)) > \mu(b)$}

= {(a, b) such that $T_R^{\mu}(a, b) > \mu(b)$}

= {(a, b)∈$E_1 \times E_2$ such that $R(a, b) > J_\mu^T(a, b)$}

6. The family of t-norms of a triangular t-norm T is the set of t-norms defined as

$T_\varphi(x, y) = \varphi^{-1}(T(\varphi(x), \varphi(y))$, given any continuous, strictly increasing function φ: [0, 1] x [0, 1] → [0, 1] such that φ(0)=0 and φ(1)=1.

Let φμ: E → [0, 1] be the fuzzy set defined by φμ(a) = φ(μ(a)). Let φR: $E_1 \times E_2$ → [0, 1] be the

fuzzy relation defined by $\varphi R(a, b) = \varphi(R(a, b))$.

7. A t-norm T is an ordinal sum if there exist a finite or numerable collection of archimedean t-norms $\{T_i / i \in J\}$ and a collection of disjoint intervals $\{(a_i,b_i\}/ i \in J\}$ in $[0,1]$ such that $T(x,y)=$

$$\begin{cases} a_i + (b_i - a_i)T_i\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right) & \text{if } (x,y) \in [a_i, b_i]^2 \\ \text{Min}(x,y) & \text{otherwise} \end{cases}$$

The proofs of both following theorems can be found in [2].

Theorem 1.

Given a fuzzy relation R, a fuzzy set μ, a function φ and a t-norm T, it holds that

$(a, b) \in INC_T{}^{\varphi\mu}(R)$ if and only if $(a, b) \in INC_T{}^{\varphi\mu}(\varphi R)$.

Given a fuzzy set $\mu: E \to [0,1]$, it is defined the fuzzy relation $\mu_2: E_1 \times E_2 \to [0,1]$ as $\mu_2(a, b) = \mu(b)$.

Theorem 2

Given the t-norm minimum or any archimedean t-norm T, for all $(a, b) \in E_1 \times E_2$, the distance $1\text{-}J^T$ between a fuzzy relation R in the point (a, b) and its μ-T-conditionalized relation in (a, b) is the same that the distance $1\text{-}J^T$ between $T_R{}^{\mu}(a, b)$ and $\mu_2(a, b)$. This is, it holds that

$J^T(R(a, b), J_\mu{}^T(a, b)) = J^T(T_R{}^{\mu}(a,b), \mu_2(a,b))$.

3 THE CASE OF ORDINAL SUMS

Lemma 1

Let T be an ordinal sum. If $x, y \in [a_i, b_i]$, then $T(x, y) \in [a_i, b_i]$.

Proof

T is monotonous, $T(a_i, a_i) = a_i$, $T(b_i, b_i) = b_i$, so $a_i \le T(x, y) \le b_i$. ∎

Lemma 2

Let T be an ordinal sum defined through a collection of archimedean t-norms $\{T_i / i \in J\}$ and a collection of disjoint intervals $\{(a_i, b_i\}/ i \in J\}$. The residuated operation of T is $J^T(x, y) = \sup\{z / T(x, z) \le y) =$

$$\begin{cases} 1 & \text{if } x \le y \\ y & \text{if } x > y \text{ and } (x, y) \notin [a_i, b_i] \quad \text{for all } i \in J \\ a_i + (b_i - a_i)J^{T_i}\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right) & \text{if } x > y \text{ and } (x, y) \notin [a_i, b_i] \end{cases}$$

Proof

1) **If $x \le y$,** any residuated operation of a t-norm takes the value 1.

2) **If $x>y$** and **$(x, y) \notin [a_i, b_i]^2$,** then $J^T(x,y) = y$. This is because

2.1) if $x \notin [a_i, b_i]$, then $J^T(x, y) = \sup\{z / T(x, z) \le y\} = \sup\{z / \text{Min}(x, z) \le y\} = y$.

2.2) if $x \in [a_i, b_i]$ and $y \notin [a_i, b_i]$, then $z = J^T(x, y)$ is not in $[a_i, b_i]$, because $y<x$, so $y<a_i$ and if $z = J^T(x, y)$ would be in $[a_i, b_i]$, then, by lemma 1, $T(x, z)$ would also be in $[a_i, b_i]$ which contradicts $T(x, z) \le y$. As $z \notin [a_i, b_i]$ the t-norm must be the minimum and $J^T(x, y) = \sup\{z / T(x, z) \le y\} = \sup\{z / \text{Min}(x, z) \le y\} = y$.

3) **If $x>y$** and **$(x, y) \in [a_i, b_i]^2$,** then $J^T(x, y) \in [a_i, b_i]$. This holds because if $z = J^T(x, y)$ would not be in $[a_i, b_i]$, then:

$J^T(x, y) = \sup\{z / T(x, z) \le y\} = \sup\{z / \text{Min}(x, z) \le y\} = y$, which contradicts with $y \in [a_i, b_i]$. So, in this case:

$$J^T(x, y) = \sup\{z / T(x,z) \le y\}$$

$$= \sup\left\{z / a_i + (b_i - a_i)T_i\left(\frac{x-a_i}{b_i-a_i}, \frac{z-a_i}{b_i-a_i}\right) \le y\right\}$$

$$= \sup\left\{z / T_i\left(\frac{x-a_i}{b_i-a_i}, \frac{z-a_i}{b_i-a_i}\right) \le \frac{y-a_i}{b_i-a_i}\right\}$$

$$= \left\{z / \frac{z-a_i}{b_i-a_i} = J^{Ti}\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right)\right\}$$

$$= a_i + (b_i - a_i)\, J^{Ti}\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right). \blacksquare$$

Definition 1.

Let R be a fuzzy relation. The fuzzy relation $\mathbf{R_{Ii}}$ restricted from R to the interval $I_i=[a_i, b_i]$, is defined by

$$R_{Ii}(a, b) = \frac{R(a, b) - a_i}{b_i - a_i}.$$

Let μ be a fuzzy set. The fuzzy set μ_{Ii} restricted from μ to the interval $[a_i, b_i]$, is defined by

$$\mu_{Ii}(a)= \frac{\mu(a) - a_i}{b_i - a_i}.$$

Theorem 3.

Let R be a fuzzy relation, let μ be a fuzzy set, and let T be an ordinal sum, then

$J^T(R(a, b), J_\mu{}^T(a, b)) = J^T(T_R{}^\mu(a, b), \mu_2(a,b))$ for all (a,b) in $E_1 \times E_2$.

Proof

1) If $\mathbf{R(a, b) \le J_\mu{}^T(a, b)}$, it is, if $(T)R^\mu(a, b) \le \mu_2(a, b)$, then $J^T(R(a, b), J_\mu{}^T(a, b)) = J^T(T_R{}^\mu(a, b), \mu_2(a, b)) = 1$.

It includes the case $\mu(a) \le \mu(b)$.

On the other cases, it is supposed that

$\mathbf{R(a, b) > J_\mu{}^T(a, b)}$, it is, $\mathbf{T_R{}^\mu(a, b) > \mu_2(a, b)}$ and $\mu(a) > \mu(b)$.

2) If $\mu(b) \notin [a_i, b_i]$, then, by lemma 2, case 2.2 it holds that:

$J^T(R(a, b), J_\mu{}^T(a, b))$
$= J^T(R(a, b), J^T(\mu(a), \mu(b))$
$= J^T(R(a, b), \mu(b))$
$= \mu(b)$
$= J^T(T_R{}^\mu(a, b), \mu(b))$
$= J^T((T)R^\mu(a, b), \mu_2(a, b))$.

3) If $\mu(a), \mu(b) \in [a_i, b_i]$ and $R(a,b) \notin [a_i, b_i]$, then, by lemma 2 it holds that

$J^T(R(a, b), J_\mu{}^T(a, b))$
$= J_\mu{}^T(a, b)$
$= J^T(\mu(a), \mu(b))$ (*)
$= J^T(\mathrm{Min}(\mu(a), R(a, b)), \mu(b))$
$= J^T(T(\mu(a), R(a, b)), \mu(b))$
$= J^T(T_R{}^\mu(a, b), \mu_2(a, b))$.

(*) $T(\mu(a), R(a, b)) > \mu(b)$, so $R(a, b)) > \mu(b)$, but $R(a, b) \notin [a_i, b_i]$, so $a_i \le \mu(b) < \mu(a) \le b_i < R(a, b))$, and $\mathrm{Min}(\mu(a), R(a, b)) = \mu(a)$.

4) If $\mathbf{R(a,b)}$, $\mu(b)$ are in $[a_i, b_i]$ and $\mu(a) \notin [a_i, b_i]$, then

$J^T(R(a, b), J_\mu{}^T(a, b))$
$= J^T(R(a, b), J^T(\mu(a), \mu(b))$
$= J^T(R(a, b), \mu(b))$ (*)
$= J^T(\mathrm{Min}(\mu(a), R(a, b)), \mu(b))$
$= J^T(T(\mu(a), R(a, b)), \mu(b))$
$= J^T(T_R{}^\mu(a, b), \mu_2(a, b))$

(*) $T(\mu(a), R(a, b)) > \mu(b)$, so $R(a, b)) > \mu(b)$, and $a_i \le \mu(b) < R(a,b) \le b_i < \mu(a)$, so $\mathrm{Min}(\mu(a), R(a, b)) = R(a, b)$.

5) If $\mu(b) \in [a_i, b_i]$ and $\mu(a), \mathbf{R(a, b)}$ are not in $[a_i, b_i]$, then

$J^T(R(a, b), J_\mu{}^T(a, b))$
$= J^T(R(a, b), J^T(\mu(a), \mu(b))$
$= \mu(b)$ (*)
$= J^T(T(\mu(a), R(a, b)), \mu(b))$
$= J^T(T_R{}^\mu(a, b), \mu_2(a, b))$

(*) By lemma 1, as $\mu(a)$, $R(a, b)$ are not in $[a_i, b_i]$, then $T(\mu(a), R(a, b))$ is not in $[a_i, b_i]$.

6) If $\mu(a), \mu(b), \mathbf{R(a, b)} \in [a_i, b_i]$, then by lemma 2, case 3, it holds that:

$J^T(R(a, b), J_\mu{}^T(a, b))$

$$= J^T\left(R(a,b), a_i+(b_i-a_i)\, J^{Ti}\left(\frac{\mu(a)-a_i}{b_i-a_i}, \frac{\mu(b)-a_i}{b_i-a_i}\right)\right)$$

$$= a_i+(b_i-a_i) J^{Ti}\left(\frac{R(a,b)-a_i}{b_i-a_i}, J^{Ti}\left(\frac{\mu(a)-a_i}{b_i-a_i}, \frac{\mu(b)-a_i}{b_i-a_i}\right)\right)$$

$$= a_i+(b_i-a_i)\, J^{Ti}\left(R_{Ii}(a,b), J^{Ti}_{\mu_{Ii}}(a,b)\right) \quad (*)$$

$$= a_i+(b_i-a_i)\, J^{Ti}\left((T_i)^{\mu_{Ii}}_{R_{Ii}}(a,b), \mu_{Ii}(b)\right)$$

$$= a_i+(b_i-a_i) J^{Ti}\left(T_i\left(\frac{\mu(a)-a_i}{b_i-a_i}, \frac{R(a,b)-a_i}{b_i-a_i}\right), \frac{\mu(b)-a_i}{b_i-a_i}\right)$$

$$= J^T\left(a_i+(b_i-a_i)T_i\left(\frac{\mu(a)-a_i}{b_i-a_i}, \frac{R(a,b)-a_i}{b_i-a_i}\right), \mu(b)\right) \quad (**)$$

$= J^T(T(\mu(a), R(a, b)), \mu(b))$
$= J^T(T_R{}^\mu(a, b), \mu_2(a, b))$.

(*) For archimedean t-norms it holds that $J^T(R(a, b), J_\mu{}^T(a, b)) = J^T(T_R{}^\mu(a,b), \mu_2(a,b))$, so, for any

t-norm T_i, the fuzzy relation R_{Ii} restricted to interval $[a_i, b_i]$ and the fuzzy set μ_{Ii} restricted to $[a_i, b_i]$ holds that $J^{Ti}\left(R_{Ii}(a,b), J^{Ti}_{\mu_{Ii}}(a,b)\right) = J^{Ti}((T_i)^{\mu_{Ii}}_{R_{Ii}}(a, b), \mu_{Ii}(b))$.

(**) By lemma 1, $T(\mu(a), R(a, b)) =$

$$a_i+(b_i-a_i)T_i\left(\frac{\mu(a)-a_i}{b_i-a_i}, \frac{R(a,b)-a_i}{b_i-a_i}\right) \in [a_i, b_i]. \blacksquare$$

Theorem 4.

Let R be a fuzzy relation, let μ be a fuzzy set, and let T be any continuous t-norm, then $J^T(R(a, b), J_\mu{}^T(a, b)) = J^T(T_R{}^\mu(a, b), \mu_2(a,b))$ for all (a, b) in $E_1 \times E_2$

Proof

It follows from theorems 2 and 3.

Theorem 5

Given any continuous t-norm T, for all (a, b) in $E_1 \times E_2$, the distance $1\text{-}J^T$ between a fuzzy relation R in the point (a, b) and its μ-T-conditionalized relation in (a, b) is the same that the distance $1\text{-}J^T$ between $T_R{}^\mu$ (a, b) and $\mu_2(a, b)$.

Proof

It follows from theorem 4

4 EXAMPLES

Similarly to the examples in [2], it is given the following examples:

1. When using the fuzzy T*-distance between two fuzzy relation R and R' defined by

$d_T(R, R') = \sup_{(a,b)\in E_1\times E_2} \{1\text{-}J^T(R(a, b), R'(a, b))\}$, for any continuous t-norm, the distance d_T between a fuzzy relation R and its μ-T-conditionalized fuzzy relation is equal to the distance d_T between $T(\mu(a), R(a,b))$ and $\mu_2(a, b)=\mu(b)$.

That is, for any continuous t-norm, a μ-T-unconditionality measure of fuzzy relation might be calculated as

$$\sup_{(a,b)\in E_1\times E_2} \{1\text{-}J^T(R(a, b), J_\mu{}^T(a, b))\} \text{ or as } \sup_{(a,b)\in E_1\times E_2} \{1\text{-}J^T((T)_R{}^\mu(a, b), \mu_2(a,b))\}.$$

2. For any continuos t-norm T, the order k moment of μ-T-unconditionality of a fuzzy relation might be calculated as

$$\iint_{(a,b)\in E_1\times E_2} (1\text{-}J^T(R(a, b), J_\mu{}^T(a, b)))^k \text{ or as } \iint_{(a,b)\in E_1\times E_2} (1\text{-}J^T((T)_R{}^\mu(a, b), \mu_2(a,b)))^k.$$

5 CONCLUSION

It is proved in [2] that given an archimedean t-norm T or the t-norm minimum, both ways described in [1] to measure the μ-T-conditionality property of fuzzy relations, give the same results when using a distance between fuzzy relations defined from the distance $1\text{-}J^T$. This paper extends the result for ordinal sums, and so, the result is proven for all continuous t-norms.

6 BIBLIOGRAPHY

[1] L. Garmendia. μ-T-Condicionalización de una relación difusa. Medida de μ-T-Condicionalidad. *Actas Estylf'97*. Pag. 281-286, 1997.

[2] L. Garmendia, E. Trillas, A. Salvador. Sobre "medir" la condicionalidad. *Actas Estylf'98*. Pag. 57-62, 1998.

[3] G. J. Klir, T. A. Folger. Fuzzy sets, uncertainty and information. *Prentice Hall.* 1988.

[4] E. Trillas. On Logic and Fuzzy Logic. *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, 1-2*. Pag. 107-137, 1993.

[5] E. Trillas, L.Valverde. An Inquiry into Indistinguishability Operators. *Aspects of Vagueness.* Editors H.J. Skala and others. Reidel Pubs (Dordrech). Pag. 231-256, 1984.

[6] S. Weber. Conditional measures and their applications to fuzzy sets. *Fuzzy Sets and Systems 42*. Pag. 73-85, 1991.

On the Equivalence between Distances and T-Indistinguishabilities *

Enric Trillas
Dept. of Artificial Intelligence
Technical Univ. of Madrid
28660 Boadilla del Monte
etrillas@fi.upm.es

Elena Castiñeira†
Dept. of Applied Mathematics
Technical Univ. of Madrid
28660 Boadilla del Monte
ecastineira@fi.upm.es

Ana Pradera
Dept. of Experimental Sciences
Rey Juan Carlos Univ.
28933 Móstoles
a.pradera@escet.urjc.es

Summary

This paper deals with the relationship between indistinguishability operators with respect to semigroups $\mathcal{T} = ([0,1], T, \leq; 1)$, where T is a t-norm, and generalized distances with respect to the semigroup $(\mathbb{R}^+, +, \leq; 0)$, that is, ordinary distances.

The paper is organized as follows. Section 2 establishes results for ordinary distances when considered as a particular case of generalized distances. Section 3 continues the work presented in [6] but in a more general framework that allows to obtain distances not only from Prod-indistinguishabilities but also from any T_φ-indistinguishability, being T a t-norm and $T_\varphi \in \mathcal{F}(T)$. This framework also allows to build T_φ-indistinguishabilities from distances. Finally, section 4 shows that, when T is an archimedean t-norm, the proposed framework provides a characterization of T_φ-indistinguishabilities based on the t-norm additive generator. When $\varphi = Id$, this result generalizes the Menger's Prod-indistinguishabilities characterization ([2]) and recovers the results found in [1].

Keywords: distances, indistinguishabilities, t-norms, archimedean t-norms.

1 Introduction

1.1. In 1967, E. Trillas introduced (see [3] and [5]) the concept of generalized distance as a function that assigns to each pair of elements from a given set a value in an ordered and commutative semigroup. More precisely, if $\mathcal{S} = (S, *, \leq; e)$ is such a semigroup with neutral element e, X is a non-empty set and $d : X \times X \to S$ is a function that verifies:

M1) $d(x,x) = e$
M2) $d(x,y) = d(y,x)$
M3) $d(x,z) \leq d(x,y) * d(y,z)$

for any $x, y, z \in X$, then d is called a **$\mathcal{S}$-generalized distance on X** (see [5] and [3]). Both the pseudo-distances and the distances defined on $\mathbb{R}^+$ by means of the operation $+$, the linear order $\leq$ and $e = 0$ are examples of generalized distances.

On the other hand, if $\mathcal{S} = (S, *, \leq; s)$ is a commutative ordered semigroup with a distinguished element s, a function $I : X \times X \to S$ is called an **$\mathcal{S}$-indistinguishability operator** of level s for the set X if, for any $x, y, z \in X$, it verifies (see [4] and [7]):

I1) $I(x,x) \geq s$
I2) $I(x,y) = I(y,x)$
I3) $I(x,y) * I(y,z) \leq I(x,z)$

If $S = [0,1]$, T is a t-norm and T^* its dual t-conorm, that is $T^*(x,y) = 1 - T(1-x, 1-y)$, then both $\mathcal{T} = ([0,1], T, \leq; 1)$ and $\mathcal{T}^* = ([0,1], T^*, \leq; 0)$ are commutative ordered semigroups. When I is a 1-level indistinguishability operator with respect to $\mathcal{T}$, the first and third axioms become: I1) $I(x,x) = 1$; I3) $T(I(x,y), I(y,z)) \leq I(x,z)$; and I is simply said to be a T-indistinguishability. When d is a generalized distance with respect to $\mathcal{T}^*$, the first and third distances axioms become: M1) $d(x,x) = 0$ and M3) $d(x,z) \leq T^*(d(x,y), d(y,z))$.

[6] was devoted to investigate the relations between generalized distances with respect to the semigroup $(\mathbb{R}^+, +, \leq; 0)$ (which will be from now on called ordinary distances) and Prod-indistinguishabilities. The aforementioned paper makes use of the two following families of functions: the set M_1, made of all the continuous and strictly decreasing functions $t : [0,1] \to \mathbb{R}^+ = [0, +\infty]$ such that $t(1) = 0$ and $t(x \cdot y) \leq t(x) + t(y)$; and the set M_2 of all the continuous and strictly decreasing functions $s : \mathbb{R}^+ \to [0,1]$ that verify $s(0) = 1$ and $s(x) \cdot s(y) \leq s(x+y)$. These functions allow to obtain, respectively, distances from Prod-indistinguishabilities and vice versa, and recover Menger's Prod-indistinguishabilities characterization theorem (see [2]).

* This work has been partially supported by the CICYT projects TIC96-1393-C06 and TIC98-0996-C02-02
† Corresponding author

1.2. Some basic results regarding t-norms should also be reviewed. As it is well-known, any t-norm T verifies $T \leq \text{Min}$ and, as a consequence, $T(x,x) \leq x$ for any $x \in [0,1]$; the boundary conditions $T(1,1) = 1$ and $T(0,0) = 0$ also hold. If T is a continuous function, then T belongs to one and only one of the following classes:

(1) $T(x,x) = x$ for any $x \in [0,1]$, and T is said to be *idempotent.*
(2) $T(x,x) < x$ for any $x \in (0,1)$, and T is said to be an *archimedean t-norm.*
(3) There exist x, y in $(0,1)$ such that $T(x,x) = x$, $T(y,y) < y$, and T is said to be a *non-idempotent and non-archimedean t-norm.*

Class (1) reduces to $T = \text{Min}$. Class (2) can be divided into two different subclasses:

(2.1) T has zero divisors, that is, there exist $x, y \in (0,1]$ such that $T(x,y) = 0$. In this case T is said to be a *non-strict archimedean t-norm* or a *nilpotent t-norm*;
(2.2) T has not zero divisors, that is, for any $x, y \in (0,1]$ it is $T(x,y) > 0$, and T is said to be a *strict archimedean t-norm.*

In general, if T is a t-norm, the family of T, $\mathcal{F}(T)$, is the set of all t-norms T_φ such that $T_\varphi = \varphi^{-1} \circ T \circ (\varphi \times \varphi)$, where $\varphi : [0,1] \to [0,1]$ is any strictly increasing bijection such that $\varphi(0) = 0$ and $\varphi(1) = 1$, that is, any order automorphism of the unit interval. $\mathcal{F}(T)$ is known as the T-family of t-norms.

Subclass (2.1) is made of all continuous t-norms that may be written as $T = \varphi^{-1} \circ W \circ (\varphi \times \varphi)$, where $W(x,y) = \text{Max}(0, x+y-1)$ is the Łukasiewicz t-norm. This subclass is the *Łukasiewicz family*, $\mathcal{F}(W)$.

Subclass (2.2) contains all continuous t-norms of the form $T = \varphi^{-1} \circ \text{Prod} \circ (\varphi \times \varphi)$, where $\text{Prod}(x,y) = x \cdot y$. This subclass is the *Product family*, $\mathcal{F}(\text{Prod})$.

Class (3) includes any t-norm that may be written as

$$T(x,y) = \begin{cases} a_i + (b_i - a_i) T_i\left(\frac{x-a_i}{b_i-a_i}, \frac{y-a_i}{b_i-a_i}\right) & (x,y) \in [a_i, b_i]^2 \\ \text{Min}(x,y) & \text{otherwise} \end{cases}$$

where $\{T_i, i \in J\}$ is a countable collection of archimedean t-norms and $(a_i, b_i) \subset [0,1]$ are disjoint intervals. Such continuous t-norms are called ordinal sums of archimedean t-norms or, for short, *ordinal sums.*

Obviously, $\{\text{Min}\} \cup \mathcal{F}(W) \cup \mathcal{F}(\text{Prod})$ is the set of all continuous t-norms that are not ordinal sums. Archimedean t-norms, that is, those that belong to $\mathcal{F}(W) \cup \mathcal{F}(\text{Prod})$, were characterized by the Aczél-Ling theorem: If T is a continuous binary operation on the unit interval, then T is an archimedean t-norm if and only if there exists a continuous and strictly decreasing function $f : [0,1] \to \mathbb{R}^+$ with $f(1) = 0$, called an *additive generator* of T, such that $T(x,y) = f^{(-1)}(f(x)+f(y))$ for any x, y in $[0,1]$, where $f^{(-1)}$ is the pseudo-inverse of f, defined by

$$f^{(-1)}(x) = \begin{cases} f^{-1}(x) & \text{if } x \in [0, f(0)] \\ 0 & \text{if } x \in (f(0), +\infty) \end{cases}$$

2 Indistinguishabilities as generalized distances

When I is a T-indistinguishability on X and $d = 1 - I$, the following characterization for T-indistinguishabilities was established in [4]:

Theorem 2.1 $I : X \times X \to [0,1]$ is a T-indistinguishability if and only if $d = 1 - I$ is a T^*-generalized distance.

The following equivalence is interesting in order to upgrade the above result to families of t-norms:

Lemma 2.2 Let T be a t-norm and $T_\varphi \in \mathcal{F}(T)$. Then $I : X \times X \to [0,1]$ is a T_φ-indistinguishability if and only if $\varphi \circ I$ is a T-indistinguishability.

Proof. Trivial, since φ is an order automorphism. ■

This equivalence allows to easily extend the scope of theorem 2.1 to any t-norm expressed by means of an order automorphism:

Theorem 2.3 Let T be a t-norm and $T_\varphi \in \mathcal{F}(T)$. A function $I : X \times X \to [0,1]$ is a T_φ-indistinguishability if and only if $d = 1 - \varphi \circ I$ is a T^*-generalized distance.

In the case of ordinary distances, last theorem's necessary condition remains valid for t-norms which are above Lukasiewicz t-norm:

Corollary 2.4 If $T_\varphi \in \mathcal{F}(T)$, T is a t-norm such that $T \geq W$ and $I : X \times X \to [0,1]$ is a T_φ-indistinguishability, then $d = 1 - \varphi \circ I$ is a distance on X with values in $[0,1]$.

Proof. I is a T_φ-indistinguishability, so $d = 1 - \varphi \circ I$ is a T^*-generalized distance. In addition, $T \geq W$ if and only if $T^* \leq W^*$, and $W^* = \text{Min}(1, \text{Sum}) \leq \text{Sum}$, where Sum represents from now on the operation $\text{Sum}(x,y) = x+y$ in $\mathbb{R}$. ■ (see [1] for a different proof).

Last corollary may be applied, among others, to any t-norm $T_\varphi \in \mathcal{F}(T)$ such that T is both a t-norm and a 2-copula (see [3]), in particular to any $T_\varphi \in \mathcal{F}(\text{Prod}) \cup \mathcal{F}(W)$. Note also that when $T = \text{Min}$, then if I is a Min-indistinguishability, $d = 1 - \varphi \circ I$ is a distance for any order automorphism φ , since $\mathcal{F}(\text{Min}) = \{\text{Min}\}$.

Nevertheless, the reciprocal of corollary 2.4 is not valid for $T = \text{Min}$ nor for $T = \text{Prod}$ (it is true when $T = W$, as it will be shown in section 4, because $1 - id$ is an additive generator of W). Let us consider the distance on $\mathbb{R}$ defined as $d(x,y) = \frac{|x-y|}{1+|x-y|}$ for any $x, y \in \mathbb{R}$. Then values $x = 0$, $y = 1/2$ and $z = 1$ prove that $I = 1 - d$ is not a Min-indistinguishability. For the product t-norm, it suffices to choose the euclidean distance on $X = [0,1]$, $d(x,y) = |x-y|$ for any x, y in $[0,1]$. In this case $I = \varphi^{-1} \circ (1-d)$ is not a T_φ-indistinguishability for any $T_\varphi \in \mathcal{F}(\text{Prod})$, and thus it is neither a Min-indistinguishability for any φ (take for example values $x = 0$, $y = 1/2$ and $z = 1$).

When corollary 2.4 is restricted to archimedean t-norms, the same result may be equivalently expressed by means of the t-norm's additive generator:

Corollary 2.5 Let T be an archimedean t-norm whose additive generator f verifies $f^{-1}(x) + f^{-1}(y) \leq 1 + f^{(-1)}(x+y)$ for any $x, y \in [0, f(0)]$. If $I : X \times X \to [0,1]$ is a T_φ-indistinguishability, $T_\varphi \in \mathcal{F}(T)$, then $d = 1 - \varphi \circ I$ is a distance on X with values in $[0,1]$.

Proof. The corollary's inequality is equivalent to $(1 - id) \circ f^{(-1)}(x+y) \leq (1-id) \circ f^{-1}(x) + (1-id) \circ f^{-1}(y)$, so $(1 - id) \circ f^{(-1)}$ is sub-additive in $[0, f(0)]$. Then, as it is well known (see [3]), it is $W \leq T$, since $1 - id$ is an additive generator of W. As a consequence, applying corollary 2.4, $d = 1 - \varphi \circ I$ is a distance. ■ (see [1] for a different proof).

Examples of functions that verify last corollary's condition are $f_1(x) = 1 - \sqrt{x}$ and $f_2(x) = (1-x)^2$.

3 More on distances and T_φ-indistinguishabilities

Let T be a continuous t-norm. Let $M_1(T, \text{Sum})$ be the set of functions $t : [0,1] \to \mathbb{R}^+ = [0, +\infty]$ which are continuous, strictly decreasing, and verify $t(1) = 0$ and $t \circ T \leq \text{Sum} \circ (t \times t)$. Let $M_2(T, \text{Sum})$ be the set of functions $s : \mathbb{R}^+ \to [0,1]$ which are continuous, strictly decreasing in $\mathbb{R}^+$ (or in an interval $[0, a]$, in which case it is $s(x) = 0$ for any $x \geq a$), verifying $s(0) = 1$ and $T \circ (s \times s) \leq s \circ \text{Sum}$. Obviously, the sets M_1 and M_2, introduced in [6] and reviewed in this paper's introduction, correspond, respectively, to $M_1(\text{Prod}, \text{Sum})$ and to the $M_2(\text{Prod}, \text{Sum})$ subset made of functions that are strictly decreasing in $\mathbb{R}^+$. In addition, if T_1 and T_2 are continuous t-norms such that $T_1 \leq T_2$, then $M_1(T_1, \text{Sum}) \subseteq M_1(T_2, \text{Sum})$, so $M_1(\text{Min}, \text{Sum})$ is the biggest set of this kind. In fact, it is $M_1(\text{Min}, \text{Sum}) = \{t : [0,1] \to \mathbb{R}^+; t \text{ continuous, strictly decreasing, } t(1) = 0\}$. On the contrary, if $T_1 \leq T_2$ then $M_2(T_2, \text{Sum}) \subseteq M_2(T_1, \text{Sum})$, being $M_2(\text{Min}, \text{Sum})$ the smallest one. It is even $M_2(\text{Min}, \text{Sum}) = \emptyset$, and the same happens for any ordinal sum, since the minimum operator is used to build this class of t-norms.

Next theorem shows how functions in $M_1(T, \text{Sum})$ allow to obtain distances from T_φ-indistinguishabilities.

Theorem 3.1 Let T be a continuous t-norm, $T_\varphi \in \mathcal{F}(T)$ and $t \in M_1(T, \text{Sum})$. If $I : X \times X \to [0,1]$ is a T_φ-indistinguishability on X, then $d = t \circ \varphi \circ I$ is a distance on X with values in $\mathbb{R}^+$.

Proof. It is easily proved using indistinguishabilities properties and $M_1(T, \text{Sum})$ definition. ■

As $M_1 = M_1(\text{Prod}, \text{Sum})$, the method to construct distances from Prod-indistinguishabilities that was obtained in [6] is recovered with last theorem, taking $T = \text{Prod}$ and $\varphi = id$.

Examples of functions in $M_1(\text{Prod}, \text{Sum})$ are $t_0(x) = 1 - x$, $t_1(x) = -\log x$, $t_2(x) = \sqrt{-\log x}$ and $t_3(x) = \frac{\log x}{\log x - 1}$ (see [6]). Then, if I is a T_φ-indistinguishability on X, with $T_\varphi \in \mathcal{F}(\text{Prod})$, then $d_0 = 1 - \varphi \circ I$, $d_1 = -\log(\varphi \circ I)$, $d_2 = \sqrt{-\log(\varphi \circ I)}$ and $d_3 = \frac{\log(\varphi \circ I)}{\log(\varphi \circ I) - 1}$ are distances on X with values in $\mathbb{R}^+$.

Another example is given by the function $t(x) = 1 - x^2$, which belongs to $M_1(W, \text{Sum})$.

The reciprocal of theorem 3.1 is, in general, not true. It is neither true for $\mathcal{F}(\text{Prod})$ nor for $\mathcal{F}(W)$, and, as a consequence, nor for $T = \text{Min}$. For $\mathcal{F}(\text{Prod})$, it is sufficient to consider function $t(x) = 1 - x$ with the same example that was given for corollary 2.4. For $\mathcal{F}(W)$, if $t(x) = 1 - x^2$ and d is the euclidean distance on $[0,1]$, then $I = \varphi^{-1} \circ t^{(-1)} \circ d$ is not a T_φ-indistinguishability for any T_φ in $\mathcal{F}(W)$ (take for example the values $x = 0$, $y = 1/2$ and $z = 1$).

In a similar way as for last theorem, next result allows to obtain T_φ-indistinguishabilities from distances, but now functions in $M_2(T, \text{Sum})$ are used. Note that, as $M_2(T, \text{Sum})$ is empty both for $T = \text{Min}$ and for any ordinal sum, this result may only be applied to archimedean t-norms.

Theorem 3.2 Let T be an archimedean t-norm, $T_\varphi \in \mathcal{F}(T)$ and $s \in M_2(T, \text{Sum})$. If $d : X \times X \to \mathbb{R}^+$ is a distance, then $I = \varphi^{-1} \circ s \circ d$ is a T_φ-indistinguishability on X.

Proof. It is easily proved from distance's properties and $M_2(T, \text{Sum})$ definition. ■

Some examples of T_φ-indistinguishabilities, with $T_\varphi \in \mathcal{F}(\text{Prod})$, are $I_0(x,y) = \varphi^{-1}\left(\frac{1}{1+d(x,y)}\right)$, $I_1(x,y) = \varphi^{-1}\left(e^{-d(x,y)}\right)$, $I_2(x,y) = \varphi^{-1} \circ \exp\left(-\sqrt{d(x,y)}\right)$ and

$I_3(x,y) = \varphi^{-1} \circ \exp\left(-\frac{d(x,y)}{d(x,y)+1}\right)$, since $s_0(x) = \frac{1}{1+x}$, $s_1(x) = e^{-x}$, $s_2(x) = e^{-\sqrt{x}}$ and $s_3(x) = e^{-\frac{x}{1+x}}$ belong to $M_2(\text{Prod}, \text{Sum})$ (see [6]). These examples are also T_φ-indistinguishabilities for $T_\varphi \in \mathcal{F}(W)$, since $M_2(\text{Prod}, \text{Sum}) \subset M_2(W, \text{Sum})$.

The reciprocal of theorem 3.2 does not hold. To prove it, let us consider $I(x,y) = e^{-|x-y|}$, which is a Prod-indistinguishability on $\mathbb{R}$, and $s(x) = \frac{1}{1+x}$, which is a function belonging to $M_2(\text{Prod}, \text{Sum})$; function $d(x,y) = s^{-1} \circ I(x,y) = \frac{1-I(x,y)}{I(x,y)}$ is not a distance on $\mathbb{R}$: values $x = 0$, $y = 1$ and $z = 2$ allow to prove it.

4 Another characterization for T-indistinguishabilities

In the previous section, mechanisms to obtain distances from T_φ-indistinguishabilities and reciprocally were stated. The conjunction of theorems 3.1 and 3.2 trivially provides the following result, which characterizes T_φ-indistinguishabilities when T is an archimedean t-norm.

Theorem 4.1 Let T be an archimedean t-norm, $T_\varphi \in \mathcal{F}(T)$ and t a function in $M_1(T, \text{Sum})$ such that $t^{(-1)} \in M_2(T, \text{Sum})$. Then $I : X \times X \to [0,1]$ is a T_φ-indistinguishability on X if and only if $d = t \circ \varphi \circ I$ is a distance on X with values in $\mathbb{R}^+$.

In addition, it can be proved that function t in theorem 4.1 is an additive generator of the t-norm T:

Theorem 4.2 Let T be an archimedean t-norm. A function t in $M_1(T, \text{Sum})$ is such that $t^{(-1)} \in M_2(T, \text{Sum})$ if and only if t is an additive generator of T.

Proof. From $t \in M_1(T, \text{Sum})$ and $t^{(-1)} \in M_2(T, \text{Sum})$, since, in addition, it is $t^{(-1)} \circ t = Id_{[0,1]}$, it follows $T(x,y) = t^{(-1)} \circ t \circ T(x,y) \geq t^{(-1)}(t(x) + t(y)) \geq T(t^{(-1)} \circ t(x), t^{(-1)} \circ t(y)) = T(x,y)$, that is, t is an additive generator for T. Reciprocally, $t \in M_1(T, \text{Sum})$ since $t \circ t^{(-1)}(x) \leq x$ holds for any $x \in \mathbb{R}^+$, and $t^{(-1)} \in M_2(T, \text{Sum})$ is true because $t^{(-1)}[t \circ t^{(-1)}(x) + t \circ t^{(-1)}(y)] = t^{(-1)}(x+y)$. ■

Last two theorems show that T_φ-indistinguishabilities, when $T_\varphi \in \mathcal{F}(T)$ and T is an archimedean t-norm, may be characterized by means of the t-norm's additive generator:

Theorem 4.3 Let T be an archimedean t-norm with additive generator f and $T_\varphi \in \mathcal{F}(T)$. Then $I : X \times X \to [0,1]$ is a T_φ-indistinguishability if and only if $d = f \circ \varphi \circ I$ is a distance on X with values in $\mathbb{R}^+$.

If $\varphi = Id$, this theorem summarizes the results previously obtained in [1] while directly working with additive generators. When particularized to the two subclasses of archimedean t-norms, it follows:

Theorem 4.4 If $T_\varphi \in \mathcal{F}(\text{Prod})$, $I : X \times X \to [0,1]$ is a T_φ-indistinguishability if and only if $d(x,y) = -p \log(\varphi \circ I(x,y))$, $p \in (0, +\infty)$, is a distance in X. If $T_\varphi \in \mathcal{F}(W)$, $I : X \times X \to [0,1]$ is a T_φ-indistinguishability if and only if $d = 1 - \varphi \circ I$ is a distance on X.

Note that in the case $\varphi = Id$, Menger's Prod-indistinguishabilities theorem (see [2]) is obtained.

5 Conclusions

This paper provides a framework that relates ordinary distances with indistinguishability operators defined with respect to parametrized families of t-norms, $T_\varphi \in \mathcal{F}(T)$, where T is a continuous t-norm. In this context, distances may be obtained from T_φ-indistinguishabilities and reciprocally. In addition, when T is an archimedean t-norm, a characterization of T_φ-indistinguishabilities is given by means of the t-norm's additive generator.

Acknowledgments Authors wish to thank an anonymous referee whose comments and informations helped them to improve this paper.

References

[1] B. De Baets and R. Mesiar, 1997, "Pseudo-metrics and T-equivalences". *J. Fuzzy Math*, Vol. 5, No. 2, pp. 471-481.

[2] K. Menger, 1951, "Probabilistic Theories of Relations", *Proc. Nat. Acad. Sciences USA*, 37, pp. 178-180.

[3] B. Schweizer and A. Sklar, 1983, *Probabilistic Metric Spaces*. North-Holland.

[4] E. Trillas, 1981, "Assaig sobre les relacions d'indistingibilitat", Actes Congrés Català de Lògica, pp. 51-59 (in Catalan).

[5] E. Trillas and C. Alsina, 1979, *Introducción a los espacios métricos generalizados*. Serie Universitaria 49, Fund. Juan March. (in Spanish).

[6] E. Trillas, S. Cubillo and E. Castiñeira, 1998, "Menger and Ovchinnikov on Indistinguishabilities, Revisited". *Proc. ESTYLF'98* , pp. 53-56.

[7] E. Trillas and L. Valverde, 1984, "An Inquiry into Indistinguishability Operators". *Aspects of Vagueness (Eds. H. Skala, S. Termini and E. Trillas), Kluwer*, pp. 231-256.

A NEW SEMANTICS FOR THE DIVISION OF FUZZY RELATIONS IN RELATIONAL DATABASES

Patrick Bosc
IRISA/ENSSAT
Technopole Anticipa BP 447
22305 Lannion Cedex France
bosc@enssat.fr

Olivier Pivert
IRISA/ENSSAT
Technopole Anticipa BP 447
22305 Lannion Cedex France
pivert@enssat.fr

Summary

Two principal semantics have been assigned to the division of fuzzy relations. In the first one, the weights of the divisor are seen as thresholds and in the other, they act as grades of importance. This paper discusses a third approach in which the weights play the role of ideal values. In this context, the notion of a graded inclusion between (fuzzy) sets moves to that of a graded equality. It is shown that the desired behavior can be reached using an appropriate fuzzy implication.

Keywords: Database systems, Fuzzy relations, Division, Graded set inclusion, Graded set equality, Fuzzy implications.

1 INTRODUCTION

In several retrieval applications, the user need is to search for elements of the system which are connected (in a certain sense) with a given set of items. In a relational setting, this need can be expressed thanks to a particular operator, known as the division of relations. In the context of a database that contains the relations PARTS(#part, part-name, price) and SALES(#store, #part, quantity), a typical query calling on a division is: "retrieve the stores that have sold (at least) all the parts priced under $40 in a quantity over 50". The result is the set of stores (#store) that are connected (in SALES) with at least all parts whose price is less than $40 with a quantity greater than 50.

It has been shown that the search activity undertaken in information retrieval where users are looking for those documents indexed by a given set of keywords pertains to the same category. This is also the case in the engineering system described in [4] where the user (a software developer) is looking for software components having a given set of characteristics. In all these cases, the set inclusion is the key tool to model the problem in which a set of desired features plays the role of a model. Then, it seems relevant to consider situations where weights are introduced in order to refine the notions of presence/absence of features. For instance, in the information retrieval framework such weights can be associated with the keywords. In doing so, any document is indexed with a set of terms (acting as topics), each of which has a weight reflecting the intensity of the topic in the document. Similarly, a user can put weights in his query. Such an approach is suggested in [1] where fuzzy sets are used and weights take their values in the unit interval.

More generally, when fuzzy set theory is used to model weights, a possible approach is to turn the classical division (and the underlying inclusion) into a graded one whose arguments are two fuzzy relations. This view is advocated in [2, 3, 4] and it turns out that different semantics are reached depending on the nature of the inclusion taken into account (implication or cardinality-based) and the role played by the weights attached to the divisor relation (importance or thresholds). In this paper, the idea is to investigate a new role for these weights when they act as "ideal values".

In section 2, the notion of a graded inclusion between fuzzy sets and that of the division of fuzzy relations are briefly recalled. Section 3 is devoted to the introduction of a new meaning of grades attached to elements appearing in the set acting as a model. It is pointed out that this view is connected with the idea of a rejection set, which, indeed, could also be of interest to enrich the possibilities offered when grades are importances or thresholds. The formalization of this approach is discussed in section 4. The need for a new set comparison operator is first shown and its definition in terms of a (double) implication is suggested. The conclusion summarizes the main points and contributions of the paper and opens some tracks for future research.

2 PREVIOUS SEMANTICS OF THE DIVISION OF FUZZY RELATIONS

As mentioned before, the division of two regular relations R(A, X) by S(A) where A and X represent sets of attributes whose underlying domains are finite, denoted

by $div_{A \div A}(R, S)$, aims at determining the X-values connected in R with (at least) all the A-values appearing in S.

The extension of the relational division consists in the definition of this operation when fuzzy relations are involved. An example is the query "find the stores which have sold a moderate number of all medium-priced parts". Formally, the usual division operation is defined as:

$$x \in div_{A \div A}(R, S) \Leftrightarrow S \subseteq \Gamma^{1}(x) \qquad (1)$$

where $\Gamma^{1}(x) = \{a \mid (x, a) \in R\}$, and a natural extension, as suggested in [5] and [3], stems from formula 1, where the usual set inclusion is changed into a gradual inclusion of the fuzzy relation S in $\Gamma^{-1}(x)$(the fuzzy set of A-values connected with x in R). In this context, inclusion degrees based on fuzzy implications or fuzzy cardinalities may be considered and if $\Rightarrow_f$ (resp. *) denotes a fuzzy implication (resp. a triangular norm), the following generic definitions of the division of two fuzzy relations are obtained:

$$\mu_{div_{A \div A}(R, S)}(x) = min_{Domain(A)}(S(a) \Rightarrow_f R(x, a)) \qquad (2)$$

$$\mu_{div_{A \div A}(R, S)}(x) = (\Sigma_{Domain(A)}(S(a) * R(x, a))) / \Sigma_{Domain(A)} S(a) \qquad (3).$$

To get an interpretation (from a user point of view) of this operation (and thus of its result), the key point lies in a proper choice of the implication (resp. the triangular norm). If a Gödel-like implication ($a \Rightarrow_G b = 1$ if $a \leq b$, b (Gödel) or b/a (Goguen) or 1 - (a - b) (Lukasiewicz) otherwise) is chosen in formula 2, the contribution of the tuple <a, x> is 1 provided R(a, x) is over S(a) which means that S(a) is seen as a threshold to be attained. If Dienes implication ($a \Rightarrow_D b = \max(1 - a, b)$) is used, S(a) plays the role of a level of importance. The contribution of the tuple <a, x> cannot equal 0 when S(a) is not completely important (strictly less than 1). Similarly, in formula 3, S(a) plays a role of threshold if the triangular norm is min and that of a level of importance for the product. Consequently, the interpretation of the division depends strongly on the type of inclusion degree and the role assigned to the weights in S-tuples, which is strongly tied to the implication or the triangular norm.

Example 1. Let us consider the query: "find the stores which have sold a moderate number of all medium-priced parts" addressed to the relations PARTS and SALES whose schema is given before. With the extensions displayed in tables 1 and 2, the results depending on the interpretation chosen by the user are:

for s1: $\min(1 \Rightarrow_{Gö} .8, .4 \Rightarrow_{Gö} .2, .6 \Rightarrow_{Gö} 1) = .2$;
$\min(1 \Rightarrow_{Gg} .8, .4 \Rightarrow_{Gg} .2, .6 \Rightarrow_{Gg} 1) = .5$;
$\min(1 \Rightarrow_{Lu} .8, .4 \Rightarrow_{Lu} .2, .6 \Rightarrow_{Lu} 1) = 8$;
$\min(1 \Rightarrow_{D} .8, .4 \Rightarrow_{D} .2, .6 \Rightarrow_{D} 1) = .6$;
$(\min(1,. 8) + \min(.4,. 2) + \min(.6, 1)) / (1 + .4 + .6) = .8$;
$(1 * .8) + (.4 * .2) + (.6 * 1) / (1 + .4 + .6) = .74$;

for s2: $\min(1 \Rightarrow_{Gö} .5, .4 \Rightarrow_{Gö} 0, .6 \Rightarrow_{Gö} 0) = 0$;
$\min(1 \Rightarrow_{Gg} .5, .4 \Rightarrow_{Gg} 0, .6 \Rightarrow_{Gg} 0) = 0$;
$\min(1 \Rightarrow_{Lu} .5, .4 \Rightarrow_{Lu} 0, .6 \Rightarrow_{Lu} 0) = .4$;
$\min(1 \Rightarrow_{D} .5, .4 \Rightarrow_{D} 0, .6 \Rightarrow_{D} 0) = .4$;
$(\min(1, .5) + \min(.4, 0) + \min(.6, 0)) / (1 + .4 + .6) = .25$;
$(1 * .5) + (.4 * 0) + (.6 * 0) / (1 + .4 + .6) = .25$ ♦

Table 1: Sample extension for the relation SALES

SALES	#store	#part	quantity	moderate
	s1	p1	30	.8
	s1	p2	99	.2
	s1	p3	40	1
	s2	p1	85	.5

Table 2: Sample extension for the relation PARTS

PARTS	#part	price	medium
	p1	100	1
	p2	300	.4
	p3	250	.6

3 ANOTHER ROLE OF THE GRADES IN THE DIVISOR

If the forementioned roles assigned to grades of the divisor are useful in some fields such as information or software component retrieval, another need appears in these domains. More precisely, the user may be interested in grading the elements of the divisor with weights that do not have a semantics of thresholds, but that act as "ideal values" (objectives to be attained as exactly as possible).

Let us illustrate this with an example in the context of information retrieval. A user is supposed to look for documents having a given "profile", i.e., addressing topic t1 at level l1, topic t2 at level l2 and so on. As with the two other interpretations (threshold and importance), it is assumed that the user knows the indexing language and the way levels (valued in the unit interval) are assigned during the indexing stage. For instance, a user interested in developing a database application with Java will choose the profile {<"databases", 1>; <"application development", 0.6>; <"Java", 0.8>}. With a threshold semantics, any document where the keyword "databases" (resp. "application development", ."Java") is assigned a grade greater than (or equal to) 1 (resp. 0.6, 0.8) is totally satisfactory and this is not the desired behavior, since when the desired "ideal" value of the profile is exceeded, the satisfaction should not be complete.

The special case where the ideal value specified for a grade is 0 deserves some particular attention. A specified zero grade means that "ideally" the associated element should be absent from the dividend. On the other hand, the elements which were not specified by the user, i.e., those which do not explicitely appear in the dividend relation S, also belong to S to a same degree 0 and thus will be considered undesirable as well. As a matter of

fact, there is no way, using a single fuzzy set, to express different meanings for the zero in order to distinguish between the notions of undesirable elements and indifferent ones. For that purpose, two sets have to be provided: a fuzzy set S^+ of desired elements whose attached grade is positive and a crisp set S^0 of elements that are undesirable. The elements of Domain(A) that do not belong to the support of $S^+ \cup S^0$ are considered "indifferent" and they should not affect the result. In doing so, an element of S^0 will be treated as a limit case of an element of S^+ when the grade moves from a "small" value to zero and the behavior of the system is continuous, i.e., there is not a big difference in the result for these two situations.

It is worth noticing that, with the two other semantics (threshold or importance), the situation is quite different: the elements which do not appear in the divisor are all considered indifferent (cf. formulas 2 and 3). However, let us remark that a set of rejected elements could also be incorporated when importance or threshold semantics is chosen. In those cases, this set could be graded in order to express the effect in terms of importance or threshold of the (more or less) high presence of an undesired element.

Clearly, the notion of an "ideal" value played by a grade calls on a sort of resemblance for the matching mechanism (whereas the threshold semantics is more a matter of attainment). The operation (denoted f) used for this purpose must satisfy (at least) the following two properties:

- $f(a, a) = 1 \ \forall a \in [0, 1]$,
- $f(a, a + \alpha) = f(a, a - \alpha) \ \forall a, (a + \alpha), (a - \alpha) \in [0, 1]$.

The second expression states that the satisfaction is the same when a same quantity (α) is exceeding or lacking the "ideal" value a.

4 IDEAL VALUES AND THE DIVISION

In this section, a formalization of the behavior described before is proposed. More precisely, the principle of the approach is to start with formula 1 and to extend it in the spirit of what has been done for thresholds and importances.

First, let us remark that formula 1 can be rewritten as:

$$x \in div_{A \div A}(R, S) \Leftrightarrow S = (S \cap \Gamma^{-1}(x)) \qquad (4)$$

where the intersection used in the right-hand side is aimed at the removal of the values of domain(A) that are "indifferent". From this, one can move to the case where the relations are fuzzy (graded) ones. Here, S is indeed made of two sets: S^+ the fuzzy set of (more or less) desired elements and S^0 the crisp set of undesirable elements. R being a fuzzy relation, $\Gamma^{-1}(x)$ becomes:

$$\mu_{\Gamma^{-1}(x)}(a) = \mu_R(x, a).$$

For the sake of clarity, it is better to give an expression of the division in the spirit of formula (4) where the two components of S are treated separately, which yields:

$$\mu_{div_{A \div A}(R, S)}(x) = min(\mu_{\approx}(support(S^+) \cap \Gamma^{-1}(x), S^+), \mu_{\approx}(S^0 \cap \Gamma^{-1}(x), \varnothing) \qquad (5)$$

where $\approx$ stands for a graded equality between fuzzy sets. The first part of the formula is a direct adaptation from (4) which deals with the desired elements (S^+) and the second part expresses that the ideal grade of any undesirable element (specified in S^0) appearing in the dividend (R) is 0.

A rather natural way for defining the graded equality between fuzzy sets appearing in formula (5) is to generalize the usual equality seen as a double inclusion:

$$\mu_{\approx}(E, F) = min(\mu_{gr\text{-}inc}(E, F), \mu_{gr\text{-}inc}(F, E))$$

where gr-inc (- , -) denotes a graded inclusion between (fuzzy) sets. Here, the two lines used in formulas 2 and 3 for the inclusion (implication or cardinality-based inclusion) can be investigated in the perspective of the satisfaction of the properties mentioned at the end of section 3 (in particular, the symmetric behavior in case of excess or default). Different definitions of the inclusion are now reviewed.

If the grade of inclusion (of E in F) taken into account is based on the ratio between the cardinality of ($E \cap F$) and E:

$$\mu_{gr\text{-}inc}(E, F) = (\Sigma_X \mu_{E \cap F}(x)) / \Sigma_X \mu_E(x)$$

it is easy to prove that, in general, the desired behavior is not reached as shown in example 2.

Example 2. Let us take three fuzzy sets E, F and G defined as follows:

$E = \{.1/a + .7/b\}$, $F = \{.1/a + .5/b\}$, $G = \{.1/a + .9/b\}$.

The expected behavior should lead to the validity of the equality: $\mu_{\approx}(E, F) = \mu_{\approx}(E, G)$. Let us compute these values from those of the various inclusion degrees involved, assuming the minimum for the intersection:

$\mu_{gr\text{-}inc}(E, F) = 0.75$, $\mu_{gr\text{-}inc}(F, E) = 1$, $\mu_{\approx}(E, F) = 0.75$,
$\mu_{gr\text{-}inc}(E, G) = 1$, $\mu_{gr\text{-}inc}(G, E) = 0.8$, $\mu_{\approx}(E, G) = 0.8$ ♦

Let us now look at (fuzzy) implication-based grades of inclusion:

$$\mu_{gr\text{-}inc}(E, F) = min_{x \in X} \mu_E(x) \Rightarrow_f \mu_F(x)$$

where $\Rightarrow_f$ is a fuzzy implication. Dienes ($a \Rightarrow_D b = \max(1 - a, b)$), Gödel ($a \Rightarrow_{Gö} b = 1$ if $a \leq b$, b otherwise) or Goguen ($a \Rightarrow_{Gg} b = 1$ if $a \leq b$, b/a otherwise) have a result which is not a linear combination of the inputs and then, they are by nature unable to reflect the desired behavior as illustrated in example 3 hereafter.

Example 3. Let us take the same fuzzy sets E, F anf G as those of example 2. The computation of the equality degree between E and F on the one hand and E and G on the other yields:

- with Dienes implication:

$\mu_{gr\text{-}inc}(E, F) = 0.5$, $\mu_{gr\text{-}inc}(F, E) = 0.7$, $\mu_{\approx}(E, F) = 0.5$,
$\mu_{gr\text{-}inc}(E, G) = 0.9$, $\mu_{gr\text{-}inc}(G, E) = 0.7$, $\mu_{\approx}(E, G) = 0.7$

- with Gödel implication:

$\mu_{gr\text{-}inc}(E, F) = 0.5$, $\mu_{gr\text{-}inc}(F, E) = 1$, $\mu_{\approx}(E, F) = 0.5$,
$\mu_{gr\text{-}inc}(E, G) = 1$, $\mu_{gr\text{-}inc}(G, E) = 0.7$, $\mu_{\approx}(E, G) = 0.7$

- with Goguen implication:

$\mu_{gr\text{-}inc}(E, F) = 5/7$, $\mu_{gr\text{-}inc}(F, E) = 1$, $\mu_{\approx}(E, G) = 5/7$,
$\mu_{gr\text{-}inc}(E, G) = 1$, $\mu_{gr\text{-}inc}(G, E) = 7/9$, $\mu_{\approx}(E, G) = 7/9$ ♦

Let us now move to the case of Lukasiewicz implication ($a \Rightarrow_L b = 1$ if $a \leq b$, $1 - (a - b)$ otherwise) and consider the key situation where an element x has a grade a in E and $(a - \alpha)$ in F. For this element x, one has:

$$\mu_E(x) \Rightarrow_L \mu_F(x) = 1 - \alpha \text{ and } \mu_F(x) \Rightarrow_L \mu_E(x) = 1.$$

The "symmetric" situation would be when the grade of x is a in E and $(a + \alpha)$ in F, for which one gets:

$$\mu_E(x) \Rightarrow_L \mu_F(x) = 1 \text{ and } \mu_F(x) \Rightarrow_L \mu_E(x) = 1 - \alpha.$$

We observe that in both cases the "local" result tied to x is the same (namely $(1 - \alpha)$), which is the expected one.

Finally, we propose to define the division of fuzzy relations where the grades of the divisor act as "ideal" values (i.e, to be attained as closely as possible) according to formula (5) where the fuzzy equality between (fuzzy) sets is defined as:

$$\mu_{\approx}(E, F) = min(min_{x \in X}\, \mu_E(x) \Rightarrow_L \mu_F(x),\ min_{x \in X}\, \mu_F(x) \Rightarrow_L \mu_E(x)).$$

Example 4. Let us illustrate the result obtained in the context of the retrieval of documents (see section 3) with:

S^+ = {1/database, 0.7/application development, 0.8/Java}
S^0 = {C, C++}

meaning that the user is interested in documents dealing with database application development in Java, but not with C or C++. If the documents:

d1 = {0.8/database, 1/application development, 1/Java, 0.4/Pascal, 0.2/C}
d2 = {1/database, 0.4/application development, 0.7/Java, 0.6/C, 0.4/C++}

are in the archive, their matching degree with respect to the query (S^+ and S^0) are:

- for d1: $\min(\mu_{\approx}(\{0.8/\text{database}, 1/\text{application development}, 1/\text{Java}\}, \{1/\text{database}, 0.7/\text{application development}, 0.8/\text{Java}\}), \mu_{\approx}(\{0.2/C\}, \emptyset)) = \min(0.7, 0.8) = 0.7$,

- for d2: $\min(\mu_{\approx}(\{1/\text{database}, 0.4/\text{application development}, 0.7/\text{Java}\}), \{1/\text{database}, 0.7/\text{application development}, 0.8/\text{Java}\}), \mu_{\approx}(\{0.6/C, 0.4/C{+}{+}\}, \emptyset)) = \min(0.7, 0.4) = 0.4$ ♦

The approach adopted here with Lukasiewicz implication has the advantage of keeping close to the general framework suggested for extending the division to fuzzy relations, which makes use of fuzzy implications. The counterpart is the rigidity of the solution due to the general shape of the comparison which, for each element x of the divisor (including S^+ and S^0), computes:

$$min(\mu_{divisor}(x) \Rightarrow_L \mu_{dividend}(x),\ \mu_{dividend}(x) \Rightarrow_L \mu_{divisor}(x)) = 1 - abs(\mu_{divisor}(x) - \mu_{dividend}(x)).$$

Then, it could then be of interest to investigate the extent to which such a function can be parameterized while keeping the connection with a double inclusion relying on an implication(-like) operator. In doing so, one could expect in particular a maximal satisfaction when grades are close but not necessarily identical and/or a minimal satisfaction for pairs of grades differing from less than 1.

5 CONCLUSION

In this paper, a new semantics is proposed for the division of fuzzy relations. This approach is complementary to the two other interpretations suggested in previous papers. The idea is to offer a unique framework in which the three possible interpretations take place and it turns out that graded inclusion and equality between fuzzy sets are central tools. The new semantics, in which the grades of the divisor are seen as ideal values to attain as closely as possible, can be modelled using Lukasiewicz implication which offers appropriate properties. The extension of this solution in order to reach a richer range of attitudes in the framework of implication-like operators is a matter for future work.

References

[1] G. Bordogna, P. Carrara and G. Pasi (1991). Query term weights as constraints in fuzzy information retrieval. *Information Processing & Management*, 27, pp 15-26.

[2] P. Bosc, G. Bordogna and G. Pasi (1996). Fuzzy inclusion in database and information retrieval query interpretation. *Proc. 6th International Conference on Information Processing and Management of Uncertainty (IPMU'96)*, Granada (Spain), July 1996, pp 1171-1176.

[3] P. Bosc, O. Pivert and H. Prade (1997). Flexible queries in relational databases - The example of the division operator. *Theoretical Computer Science*, 171, pp 281-302.

[4] P. Bosc, E. Damiani and M-G. Fugini (1998). Dynamic service identification in a CORBA-like environment. *Proc. 1st International Workshop on Innovative Internet Information Systems (IIIS'98)*, Pisa (Italy), pp 161-175.

[5] D. Dubois and H. Prade (1996). Semantics of quotient operators in fuzzy relational databases, *Fuzzy Sets and Systems*, 78, pp 89-93.

CONTEXTOS PONDERADOS.

Ana Burusco
Dpto. de Automática y Computación
Universidad Pública de Navarra
31006 Pamplona, Spain
e-mail: burusco@si.upna.es

Ramón Fuentes González
Dpto. de Automática y Computación
Universidad Pública de Navarra
31006 Pamplona, Spain
e-mail: rfuentes@si.upna.es

Resumen

Se proponen varios métodos basados en la teoría de conceptos L-Fuzzy para el estudio de contextos que recogen la opinión de varios expertos sobre la relación existente entre dos conjuntos, a los que llamaremos de objetos y de atributos, partiendo de la premisa de que la opinión de cada experto tiene un peso o credibilidad distinto.

Palabras clave: Conceptos L-Fuzzy, Conceptos L-Fuzzy k-valuados ponderados, Métodos de análisis de datos, Procesamiento de la Información.

1 INTRODUCCION

En [3] definimos contextos M(L)-Fuzzy k-valuados como sistemas relacionales $(M_k(L), X, Y, \widetilde{\widetilde{R}})$ donde X e Y eran los conjuntos de objetos y atributos, L la cadena $\{0, 0.1, 0.2 \ldots, 1\}$ y $\widetilde{\widetilde{R}} : X \times Y \longrightarrow M_k(L)$, siendo $M_k(L)$ el conjunto de multiconjuntos de L con cardinalidad k.

Estos contextos representaban la opinión de k expertos sobre la relación existente entre los objetos X y los atributos Y. En aquella ocasión, dábamos la misma credibilidad a todos los expertos, de ahí que la imagen de nuestros conjuntos y relaciones M(L)-Fuzzy k-valuados fuesen multiconjuntos de cardinalidad k, no teniendo importancia el orden de las observaciones.

Una vez definido el contexto en el que se iba a trabajar, se daban las definiciones de operadores derivación que permitían construir el retículo de los conceptos M(L)-Fuzzy k-valuados a partir de los cuales se extraía la información del contexto correspondiente manteniendo la diferencia de opiniones mostrada por los expertos.

La situación que planteamos en este trabajo cambia en cuanto que asociamos a los k expertos unos pesos $P = (p_1, p_2, \ldots p_k)$ que representan su grado de credibilidad, así p_l representa el grado de credibilidad del experto $l-$ésimo.

Este problema no podría resolverse mediante el planteamiento propuesto en [3] ya que, en aquel caso, se trabajaba con multiconjuntos cuyas observaciones aparecían ordenadas. Si utilizásemos ahora multiconjuntos, tendríamos dificultades para saber a qué experto corresponde cada observación y cuál es por tanto su peso. En realidad, estaríamos mezclando opiniones de distintos expertos sin tener en cuenta sus pesos.

En el siguiente apartado plantearemos de forma explícita cuáles son las claves del problema a modelizar.

2 CONTEXTOS L-FUZZY K-VALUADOS PONDERADOS

Sea L una cadena completa en [0,1] y $X = \{x_1, x_2 \ldots x_n\}$ e $Y = \{y_1, y_2 \ldots y_m\}$ conjuntos no vacíos a los que llamaremos conjuntos de objetos y de atributos respectivamente. Supongamos también que k expertos opinan sobre la relación existente entre cada uno de los objetos de X y cada uno de los atributos de Y, siendo $\overset{\frown}{R} : X \times Y \longrightarrow L \times \overset{k}{\ldots} \times L$ la tabla que recoge dichas opiniones y tomando $P = (p_1, p_2, \ldots p_k)$ como pesos de los expertos, con $p_l \in L, \forall l = 1 \ldots k$, tales que $\sum p_l = 1$. Así, por ejemplo, $\overset{\frown}{R}(x_i, y_j) \in L^k$ representará la opinión de los expertos sobre la relación existente entre el objeto x_i y el atributo y_j y $\overset{\frown}{R}_l(x_i, y_j)$, la opinión del experto $l-$ésimo sobre la relación existente entre dicho objeto y dicho atributo siendo su grado de credibilidad p_l.

Habitualmente, la cadena L con la que trabajaremos en nuestros ejemplos será $L = \{0, 0.1, \ldots, 1\}$.

2.1 Definición Llamaremos a $\overset{\frown}{R}$ relación L-Fuzzy k-valuada ponderada y a la tupla $(L^k, X, Y, P, \overset{\frown}{R})$ contexto L-Fuzzy k-valuado ponderado.□

La imagen de la aplicación $\overset{\frown}{R}$, tal y como ha sido definida, es un elemento de L^k, que es un retículo completo con el orden usual:

$$(\alpha_1, \alpha_2 \ldots \alpha_k) \leq (\beta_1, \beta_2 \ldots \beta_k) \Longleftrightarrow \alpha_l \leq \beta_l, \forall l = 1 \ldots k$$

A continuación presentamos tres posibles formas de analizar y extraer información de este tipo de contextos: La primera consiste en obtener, mediante agregaciones, un contexto L-Fuzzy $(L, X, Y, \overset{\frown}{R}_{agre})$ y aplicar las técnicas descritas en [1,2] para obtener los conceptos correspondientes.

La segunda forma de abordar el problema será interpretar el contexto $(L^k, X, Y, P, \breve{R})$ como k contextos L-Fuzzy independientes, obtener sus conceptos L-Fuzzy y a partir de ellos aplicar el peso correspondiente a cada uno de los expertos. Finalmente, indicamos cómo se pueden adaptar las definiciones de los operadores derivación a este nuevo supuesto, incluyendo en las mismas los pesos que corresponden a los distintos expertos.

3 AGREGACION DE RELACIONES EN CONTEXTOS L-FUZZY K-VALUADOS PONDERADOS

Dado un contexto L-Fuzzy k-valuado ponderado $(L^k, X, Y, P, \breve{R})$, con L cadena completa en [0,1], X e Y conjuntos de objetos y atributos, $P = (p_1, p_2 \ldots, p_k)$ pesos de los expertos y $\breve{R} : X \times Y \longrightarrow L^k$, podemos considerar la familia de contextos L-Fuzzy $(L, X, Y, \breve{R}_l)$, donde $\breve{R}_l$ es la proyección l-ésima de $\breve{R}$ $(l = 1 \ldots k)$.

Una forma de consensuar la opinión de estos k expertos, teniendo en cuenta su grado de credibilidad o peso, consistirá en agregar las relaciones L-Fuzzy $\breve{R}_l$, eliminando así la subjetividad de los expertos, y obteniendo valores más democráticos y realistas.

En el marco en el que estamos trabajando proponemos que la agregación a utilizar verifique las propiedades: Simetría, unanimidad, monotonía y que los valores resultantes de la agregación se encuentren entre el mínimo y el máximo.

En esta linea, tal y como indican algunos autores que han estudiado agregaciones en entornos aplicados (Montero y Ramakrishnan&Rao [5]), hemos escogido la media ponderada, cuya trascripción al entorno en el que trabajamos permite obtener una relación agregada:

$$\breve{R}_{agre}(x_i, y_j) = \sum_l p_l \breve{R}_l(x_i, y_j) \qquad (1)$$

Si los valores resultantes de la agregación no pertenecen a la cadena L, aproximaremos al valor más cercano en L.

Una vez aplicada la agregación, $(L, X, Y, \breve{R}_{agre})$ será un contexto L-Fuzzy sobre el que podremos construir los conceptos L-Fuzzy tal y como se estudió en [1,2].

De entre todos los métodos propuestos, éste es el más sencillo de aplicar si lo que se quiere es obtener una información resumida y veraz que se derive de la tabla inicial, sin embargo, tiene el inconveniente de que se pierde la visión particular de cada uno de los expertos.

3.1 Ejemplo Supongamos que dos médicos opinan sobre la influencia de los factores $X = \{x_1, x_2\}$ en la evolución de dos enfermedades $Y = \{y_1, y_2\}$:

Tabla 1

$\breve{R}$	y_1	y_2
x_1	0.3 0.5	0.5 0.5
x_2	0.2 0.6	1 0.8

y que la credibilidad del primero de ellos es mayor que la del segundo: $p_1 = 0.7, p_2 = 0.3$.

Esta tabla representará un contexto L-Fuzzy k-valuado ponderado $(L^k, X, Y, P, \breve{R})$ con $k = 2$ y $L = \{0, 0.1, \ldots, 1\}$.

Si agregamos este contexto L-Fuzzy k-valuado ponderado según (1) (redondeando los valores para que pertenezcan a la cadena L), obtendremos la siguiente tabla:

Tabla 2

$\breve{R}agre$	y_1	y_2
x_1	0.4	0.5
x_2	0.3	0.9

de forma que al tomar el conjunto L-Fuzzy de factores (objetos):

$$\breve{A} = \{x_1/0, x_2/1\}$$

obtenemos el concepto L-Fuzzy

$$(\{x_1/0.5, x_2/0.7\}, \{y_1/0.3, y_2/0.5\})$$

de donde se desprende que *el segundo (x_2) es el factor que más influye y lo hace sobretodo en la segunda enfermedad (y_2).*

4 AGREGACION DE CONCEPTOS L-FUZZY

Tomemos ahora para cada experto las relaciones $\breve{R}_l$, con $l = 1 \ldots k$ definidas en el apartado anterior y construyamos los contextos L-Fuzzy $(L, X, Y, \breve{R}_l)$. Según la teoría desarrollada en [1,2], podemos construir para cada contexto $(L, X, Y, \breve{R}_l)$ el retículo de conceptos L-Fuzzy $\breve{\mathcal{L}}_l$ que se deriva de él.

Tal y como se detalla en [2], para construir dichos conceptos utilizaremos los operadores φ_l y ψ_l definidos para cada uno de los k contextos L-Fuzzy $(L, X, Y, \tilde{R}_l)$.

Una vez construidos estos retículos de conceptos, es posible agregar conceptos de distintos retículos teniendo en cuenta sus pesos. Lo más interesante será agregar conceptos que provengan de un mismo conjunto L-Fuzzy de partida. En este sentido, para cada conjunto L-Fuzzy $\tilde{A}$ podemos obtener los k puntos fijos que de este conjunto se derivan sin mas que aplicar cada uno de los operadores de P. y R. Cousot ([4]), $luis(\varphi_l) \circ llis(f_{l,2})$ siendo $luis(\varphi_l)(\tilde{A}) = \limsup(\tilde{A}, \varphi_l(\tilde{A}), \varphi_l^2(\tilde{A}) \ldots)$ y $llis(f_{l,2})(\tilde{A}) = \liminf(\tilde{A}, \tilde{A} \wedge \varphi_l(\tilde{A}), \tilde{A} \wedge \varphi_l(\tilde{A}) \wedge \varphi_l(\tilde{A} \wedge \varphi_l(\tilde{A})) \ldots)$. Los denotaremos $\tilde{A}_1, \tilde{A}_2, \ldots \tilde{A}_k$.

Para cada uno de éstos puntos fijos, construiremos su concepto L-Fuzzy correspondiente: $(\tilde{A}_1, \tilde{B}_1)$, $(\tilde{A}_2, \tilde{B}_2), \ldots$ $(\tilde{A}_k, \tilde{B}_k)$ siguiendo la teoría dada en [2].

Definamos ahora

$$\mu : L^X \longrightarrow L^X \times L^Y$$
$$\tilde{A} \hookrightarrow (\tilde{M}, \tilde{N})$$

donde $\tilde{M}(x) = \sum_l p_l \tilde{A}_l(x)$, y $\tilde{N}(y) = \sum_l p_l \tilde{B}_l(y)$, asociando a cada conjunto L-Fuzzy la agregación ponderada de los conceptos que se obtienen a partir de él.

Siguiendo este proceso, más laborioso que el del apartado anterior, se mantienen hasta el final las apreciaciones de los distintos expertos; sin embargo, no tenemos la seguridad de que el resultado final sea un concepto. (La agregación de conceptos no tiene por qué ser un concepto, de hecho, en el siguiente ejemplo tendremos que proceder a un redondeo obteniéndose finalmente un concepto L-Fuzzy.)

4.1 Ejemplo Si partimos del mismo contexto L-Fuzzy k-valuado ponderado que utilizamos en el apartado anterior, podemos construir las relaciones siguientes:

Tabla 3

$\tilde{R}_1$	y_1	y_2
x_1	0.3	0.5
x_2	0.2	1

Tabla 4

$\tilde{R}_2$	y_1	y_2
x_1	0.5	0.5
x_2	0.6	0.8

de forma que para el mismo conjunto de partida:

$$\tilde{A} = \{x_1/0, x_2/1\}$$

obtenemos el concepto L-Fuzzy:

$$(\{x_1/0.5, x_2/0.8\}, \{y_1/0.2, y_2/0.5\})$$

mediante el primer contexto y el concepto L-Fuzzy

$$(\{x_1/0.5, x_2/0.6\}, \{y_1/0.5, y_2/0.5\})$$

del segundo, de donde se desprende que *de la opinión del primer experto se deriva una mayor influencia del segundo factor (x_2), principalmente en la última de las enfermedades (y_2); mientras que según el segundo, dicho factor influye por igual en ambas enfermedades.*

Podemos ahora asociar al conjunto de partida $\tilde{A}$ el resultado de agregar ponderadamente ambos conceptos (p_1=0.7,p_2=0.3) tal y como aparece definido en la aplicación μ :

$$(\tilde{M}, \tilde{N}) = (\{x_1/0.5, x_2/0.7\}, \{y_1/0.3, y_2/0.5\})$$

(Hemos redondeado los resultados para que los grados de pertenencia sean valores de la cadena L.)

5 OPERADORES DERIVACION PARA CONTEXTOS L-FUZZY K-VALUADOS PONDERADOS.

Partamos ahora de un contexto L-Fuzzy k-valuado ponderado $(L^k, X, Y, P, \tilde{R})$ y tratemos de definir los operadores derivación que se adapten a esta nueva situación:

Para cada $\tilde{A} \in (L^k)^X$ y $\tilde{B} \in (L^k)^Y$ vamos a definir los conjuntos $\tilde{A}_1 \in (L^k)^Y$ y $\tilde{B}_2 \in (L^k)^X$ de la siguiente forma:

$$\tilde{A}_1(y) = \bigwedge_{x \in X} \left(\tilde{A}'(x) \bigvee (P \star \tilde{R}(x, y)) \right)$$
$$\tilde{B}_2(x) = \bigwedge_{y \in Y} \left(\tilde{B}'(y) \bigvee (P \star \tilde{R}(x, y)) \right)$$

donde los operadores $\bigvee$ y $\bigwedge$ son el supremo y el ínfimo definidos en L^k, $'$ el operador complementario usual y $P \star \tilde{R}(x, y) = (p_1 \tilde{R}_1(x, y), p_2 \tilde{R}_2(x, y), \ldots, p_k \tilde{R}_k(x, y))$.

5.1 Definición Llamaremos a los operadores 1 y 2 operadores derivación.□

5.2 Proposición Los operadores 1 y 2 son antítonos.

$$\tilde{A} \leq \tilde{C} \Longrightarrow \tilde{A}_1 \geq \tilde{C}_1, \forall \tilde{A}, \tilde{C} \in (L^k)^X$$
$$\tilde{B} \leq \tilde{D} \Longrightarrow \tilde{B}_2 \geq \tilde{D}_2, \forall \tilde{B}, \tilde{D} \in (L^k)^Y .\square$$

Podemos ahora definir los operadores φ y ψ combinando los operadores anteriores: $\varphi(\breve{A}) = \breve{A}_{12}$ and $\psi(\breve{B}) = \breve{B}_{21}$.

5.3 Proposición Los operadores φ y ψ preservan el orden.□

Como L^k es un retículo completo, podemos probar, al igual que hicimos en [1] los siguientes resultados:

Sea $\varphi : (L^k)^X \rightarrow (L^k)^X$ $t.q.$ $\varphi(\breve{A}) = \breve{A}_{12}$ y $fix(\varphi) = \{\breve{A} \in (L^k)^X / \breve{A} = \varphi(\breve{A})\}$, el conjunto de puntos fijos de φ.

Utilizando el teorema de puntos fijos de Tarski([6]), como $(L^k)^X$ y $(L^k)^Y$ son retículos completos y φ y ψ preservan el orden, podemos concluir que $\Omega = (fix(\varphi), \leq)$ y $\Sigma = (fix(\psi), \geq)$ son retículos completos, donde $\leq$ es la relación de orden en L^k y $\geq$ su opuesta.

5.4 Definición Si $\breve{M} \in fix(\varphi)$, entonces llamaremos al par $(\breve{M}, \breve{M}_1)$, concepto L-Fuzzy k-valuado ponderado del contexto $(L^k, X, Y, P, \breve{R})$.□

Es sencillo probar que los conjuntos $\breve{\mathcal{L}}_1 = \{(\breve{M}, \breve{M}_1) / \breve{M} \in fix(\varphi)\}$ y $\breve{\mathcal{L}}_2 = \{(\breve{N}_2, \breve{N}) / \breve{N} \in fix(\psi)\}$ son iguales, por tanto, utilizaremos $\breve{\mathcal{L}}$ para denotar a estos conjuntos. Además, el orden en el conjunto de los conceptos será el siguiente:

$\forall(\breve{A}, \breve{B}), (\breve{C}, \breve{D}) \in \breve{\mathcal{L}}$

$(\breve{A}, \breve{B}) \preceq (\breve{C}, \breve{D}) \Longleftrightarrow \breve{A} \leq \breve{C}$ (o equiv. $\breve{B} \geq \breve{D}$)

5.5 Teorema El conjunto $\breve{\mathcal{L}}(L^k, X, Y, P, \breve{R})$ con la relación de orden $\preceq$ es un retículo completo. Lo llamaremos retículo de conceptos L-Fuzzy k-valuado ponderado.□

La teoría de P. y R. Cousot ([4]) nos permite también obtener los conceptos que se derivan de un determinado conjunto de partida. En este sentido, para cada $\breve{A} \in (L^k)^X$, podemos aplicar el operador $luis(\varphi) \circ llis(f_2)$ definido en el apartado anterior y obtener de esta manera el punto fijo $\breve{A}^*$ que de él se deriva. A continuación, construiremos el concepto L-Fuzzy k-valuado ponderado correspondiente $(\breve{A}^*, \breve{A}_1^*)$.

Éste es el único método que permite obtener conceptos L-Fuzzy ponderados pero no agregados, de forma que siempre se mantiene la información de todos los expertos. El inconveniente que tiene es que al aplicar los pesos se deforman los datos iniciales (disminuyen proporcionalmente al peso).

5.6 Ejemplo En este caso podemos partir del conjunto L-Fuzzy k-valuado

$$\breve{A} = \{x_1/(0,0), x_2/(1,1)\}$$

obteniéndose el concepto:

$$(\{x_1/(0.4,0.8), x_2/(0.7,0.8)\},$$
$$\{y_1/(0.3,0.2), y_2/(0.6,0.2)\})$$

Aquí vemos que *teniendo en cuenta ambos expertos y sus pesos, éstos difieren básicamente en su apreciación sobre el primer factor (x_1) y la segunda enfermedad (y_2).*

6 CONCLUSIONES

En este trabajo hemos presentado tres posibles formas de analizar los contextos L-Fuzzy provenientes de los opiniones ponderadas de varios expertos. La elección de cada uno de estos métodos dependerá de la situación concreta que queramos estudiar.

En este sentido, si lo que interesa es obtener una información resumida que tenga en cuenta los distintos pesos de los expertos, aconsejaremos agregar ponderadamente primero, y construir los conceptos después (apartado 3). Si, por el contrario, interesa mantener la información de cada uno de los expertos de forma independiente hasta el final utilizaremos el método descrito en el apartado 4. (Considerar contextos L-Fuzzy independientes y agregar y ponderar los conceptos que se obtienen al final). Finalmente, cuando únicamente interese ponderar pero no agregar, trabajaremos con los conceptos L-Fuzzy ponderados (apartado 5).

REFERENCIAS

[1] A. Burusco, R. Fuentes-González, The Study of the L-Fuzzy Concept Lattice, Mathware and Soft Computing 1, No 3 (1994) 209-218.

[2] A. Burusco, R. Fuentes-González, Construction of the L-Fuzzy Concept lattice, Fuzzy Sets and Systems 97, No. 1 (1998) 109-114.

[3] A. Burusco, R. Fuentes-González, Contexts obtained from several expert opinions, Sixth IEEE International Conference on Fuzzy Systems, Barcelona, (1997), 339-342.

[4] P. Cousot, R. Cousot, Constructive versions of Tarski's fixed point theorems, Pacific Journal of Mathematics 82 (1979) 43-57.

[5] J. Fodor, M. Roubens, Fuzzy preference modelling and multicriteria decision support, Kluwer Academic Publishers, (1994).

[6] A. Tarski, A Lattice Theoretical Fixpoint Theorem and its Applications, Pathific J.Math. 5 (1955) 285-310.

UNFUZZY: Fuzzy Logic System analysis, design, simulation and implementation software

Oscar G. Duarte
Department of Electrical Engineering
Universidad Nacional de Colombia
Ciudad Universitaria Ed 453 Of.202
Bogotá . Colombia
ogduarte@ingenieria.ingsala.unal.edu.co

Gustavo Pérez
Department of Electrical Engineering
Universidad Nacional de Colombia
Ciudad Universitaria Ed 453 Of.202
Bogotá . Colombia
gustavo@ingenieria.ingsala.unal.edu.co

Abstract

In this paper we show a free distribution software tool useful for analysis, design, simulation and implementation of Fuzzy Logic Systems (FLS); we have called it UNFUZZY. It runs in operative system Windows 32bits, and helps the user in the task of FLS design through a grapical interface.The package has also two different training algorithms, and is able to generate source code in C and C++ language, so the user can implement in software the FLS. We also show in brief some of the applications we have had notice.

Keywords: Fuzzy Logic Systems, Fuzzy Controllers, Software

1 INTRODUCTION

UNFUZZY[1] is a free distribution software tool , whose main purpose is to make easier some usual tasks related with Fuzzy Logic Systems (FLS) analisis, design, simulation and implementation. It runs on Windows 32 bits, and is capable to manipulate to FLS with an internal structure as shown in figure 1. The main aim of this paper is to divulgate the UNFUZZY scope and to enumerate some of the applications that different users have reported us.

2 DESIGN OPTIONS

Fuzzy Logis System design is made using the graphical interface of UNFUZZY (see figure 2). The design options of UNFUZZY are listed below:

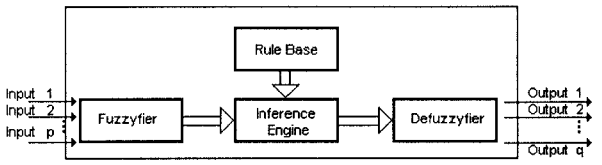


Figure 1: Fuzzy Logic System.

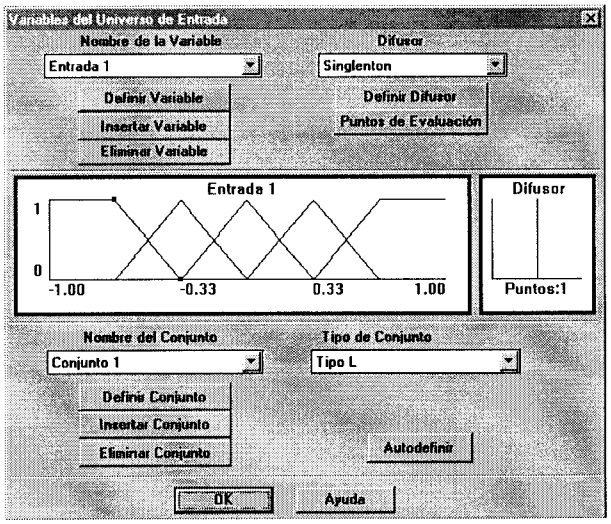


Figure 2: Graphicak interface

- Unlimitated number of input and output variables.
- Unlimitated number of labels in every linguistic variable
- 9 types of membership functions for every label
- 5 types of fuzzyfication operators for the input variables

[1]This package has been developed at the Engineering Faculty of National University of Colombia (Universidad Nacional de Colombia), and is free availabe in internet at address http://ohm.ingsala.unal.edu.co/ogduarte

- 5 types of defuzzyfication operators for the output variables
- 9 implication operators
- 9 AND operators
- 9 composition operators
- 13 agregation operators
- Rule base is inherently coherent
- In every rule, the antecedent terms may have linguistic hedges.
- Input and output variables can be auto- defined with an usual partition.
- Rule base can be auto-defined using increasing, decreasing or constant tendences.

3 FLS DESIGN USING TRAINING ALGORITHMS

UNFUZYY has two different training algorithms for automation FLS design

- Fixed Universes: it is a descriptive algorithm proposed by Wang & Mendel [13]; user can get a rule base from a set of training patterns; user must design previously the other components of the FLS.
- Variable Universes: it is an aproximative algorithm proposed by Wang [7]; user can get the labels of the input and output variables, as well as the rule base, from a set of training patterns; user must design previously the other components of the FLS.

4 ANALYSIS AND SIMULTION TOOLS

UNFUZZY has the followign analysis and simulation tools:

- It calculates the crisp outputs for a set of crisp inputs
- It makes the previous calculus using an Step by Step procedure with graphical views (see figure 3).
- It generates Input-Output tables
- It shows graphically the Input-Output relation. Graphics are bidimensional, so user can do sensibility analysis (see figure 4).
- It creates a text file with a description of the FLS.

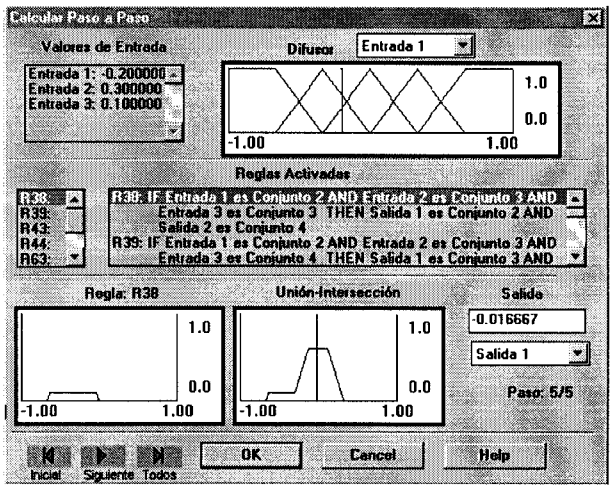


Figure 3: Step by setp analysis

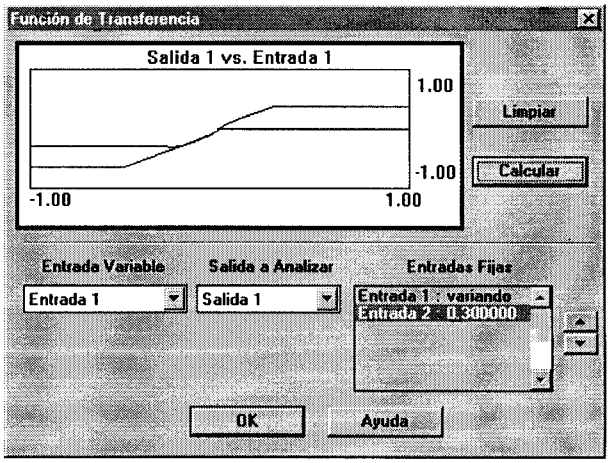


Figure 4: Transfer function

5 IMPLEMENTATION

UNFUZZY makes easy software implementation, because it generates source code; this source code can be:

- Structurated code using C language: the functions and procedures, and a brief demostrative program are written in the same file. The program shows how to use the routines.

- Object Oriented code using C++: A constructor of the FLS class and a brief demostrative program are written in the same file. The user gives the class name; this class inheritates propierties and functions from a more generic class that is specified in two different files (available with UNFUZZY). If user chooses Object Oriented code , then he/she can use the training algorithms explained in the previous section.

Structurated code is a good option for those users with a low programing level; in the other hand, Object Oriented code is very useful if the application has two or more FLS.

6 APPLICATIONS

Since september 1997, when realease 1.0 was developed, we have received some messages from the users of UNFUZZY. According to these messages, the package has been used at least in Colombia, Venezuela, Cuba, Brasil, Ecuador, Bolivia, Chile, Argentina, Mxico, Spain, Italy, Canad and USA (see figure 5).

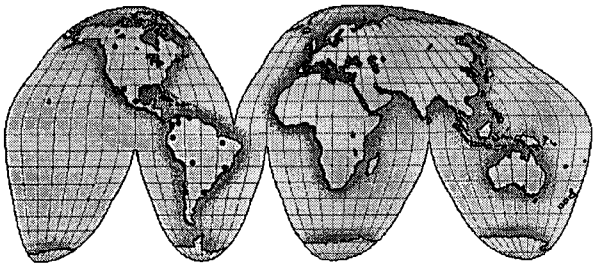


Figure 5: Impact of UNFUZZY

Some of the applications that has been reported are listed bellow:

- Controllers design:
  - Air conditioning control
  - Control to cure tobacco
  - Industrial oven control
- Fuzzy control hardware:
  - contrasting results
- Robotics:
  - Robot navigation system
- Educational porpose:
  - Didactic tool in undergraduate and postgraduate courses
  - Simulations
  - Helping resarch projects
- Decision Support Systems:
  - National Planning of Energy resources
  - Cost-benefit analysis in power plants
- Expert Systems:
  - Inventary managment system
  - Helping gas chromatografic analysis

References

[1] Alvarez, Hernán D. Control Difuso y Sistemas de Control Inteligentes. *Memorias del Segundo Congreso de la Asociación Colombiana de Automática*, Bucaramanga, Colombia, march-1997, pp 331-340

[2] Driankov, Dimiter & others. *An Introduction to Fuzzy Control* Springer Verlag, Berlin, 1993

[3] Duarte, Oscar G. UNFUZZY - Software para el análisis, diseño, simulación e implementación de Sistemas de Lgica Difusa. *M.Sc. Tesis.* Universidad Nacional de Colombia, Facultad de Ingeniería, Maestría en Automatización Industrial. 1997.

[4] Klir, George & Yuan, Bo. *Fuzzy Sets and Fuzzy Logic* Prentice Hall, New Jersey, 1995

[5] Lee, Chuen Chien. Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part I. *IEEE Transactions on Systems, Man, and Cibernetics* V 20, No. 3 March/April 1990, pp 404-418

[6] Lee, Chuen Chien. Fuzzy Logic in Control Systems: Fuzzy Logic Controller-Part II. *IEEE Transactions on Systems, Man, and Cibernetics* V 20, No. 3 March/April 1990, pp 419-435

[7] López, José F. & others. Comparacin entre sistemas difusos para el control de una banda transportadora dentro de un proceso de manufactura flexible. *Memorias del Segundo Congreso de la Asociación Colombiana de Automática*, Bucaramanga, Colombia, march-1997, pp 280-288

[8] Mendel, Jerry. Fuzzy Logic Systems for Engineering: A Tutorial. *Procceedingos of the IEEE* V 83, No. 3 March 1995, pp 345-377

[9] Meneses, Jorge E. & others, Control Fuzzy basado en microcontrolador aplicado a la operación de un válvula remota. *Memorias del Segundo Congreso de la Asociación Colombiana de Automática* Bucaramanga, Colombia, march-1997, pp 175-181

[10] Passino, Kevin, Intelligent Control. *The Control Handbook* IEEE pp 994-1001

[11] Passino, Kevin, Fuzzy Control. *The Control Handbook* IEEE pp 1001-1017

[12] Solano, Karim & others, Controladores adaptables basados en mecanismos de inferencia Difusa *Memorias del Segundo Congreso de la Asociación Colombiana de Automática*, Bucaramanga, Colombia, march de 1997, pp 104-113

[13] Wang, Li-Xin *Adaptative Fuzzy Systems and Control. Design and Stability Analysis* Prentice Hall, New Jersey, 1994

[14] Wang, Li-Xin Fuzzy Systemes are Universal Approxximators *Proceedings of the IEEE International Conference on Fuzzy Systems* San Diego, California 1992, pp 1163-1170

[15] Wang, Li-Xin & Mendel, Jerry. Back Propagation Fuzzy System as Nonlinear Dynamic Systems Identifiers *Proceedings of the IEEE International Conference on Fuzzy Systems* San Diego, California, 1992 pp 1409-1418

[16] Wang, Li-Xin & Mendel, Jerry. Fuzzy Basis Functions, Universal Approximation, ang Orthogonal Least-Squares Learning *IEEE Transactions on Neural Networks*, Vol 3, No 5, September 1992, pp 807-814

[17] Wang, Li-Xin & Mendel, Jerry. Generating Fuzzy Rules by Learning from Examples *Proceedings of the IEEE 1991 International Symposium on Intelligent Control* 13-15 August, Arlington, Virginia, U.S.A. pp 263-268

[18] Yamakawa, T. A Fuzzy Inference Engine in Nonlinear Analog mode and its applications to a fuzzy logic control *IEEE Transactions on Neural Networks*, May 1993.

Deformed Fuzzy Automata for the Text Error Correction Problem

J. Echanobe
Dpto. Matemática e Informática
Univ. Pública de Navarra
Campus de Arrosadía
31006 Pamplona, Spain
javi@we.lc.ehu.es

J. R. Garitagoitia
Dpto. Matemática e Informática
Univ. Pública de Navarra
Campus de Arrosadía
31006 Pamplona, Spain
joserra@si.unavarra.es

J. R. González de Mendívil
Dpto. Automática y Computación
Univ. Pública de Navarra
Campus de Arrosadía
31006 Pamplona, Spain
mendivil@si.unavarra.es

Summary

A fuzzy method for the text error correction problem is introduced. The method is able to handle insert, delete and change errors. Moreover, it uses the measurement level output that an Isolated Character Classifier can provide. The method is based on a Deformed System, in particular, a deformed fuzzy automaton is defined to model the possible errors in the words of the texts. Experimental results show good performance in correcting the three types of errors.

Keywords: Deformed Systems, Fuzzy automata, Contextual postprocessing, Text recognition.

1 Introduction

The automatic detection and correction of errors is an important problem in the recognition of texts. Textual errors are mainly caused during the recognition process, and they are known as edition errors: insert, delete or change errors. In text recognition systems, the error correction is in part provided by a Contextual Postprocessing (CP). Let $\omega = a_1 a_2 \dots a_m$ be an observed word which is obtained from a previous stage of the system; being the characters a_i $(1 \leq i \leq m)$ belong to an alphabet Σ. The objective of the CP is to estimate a word $\hat{\omega}$ in a set of words D (a dictionary) that is the best selection for ω, e.g., it minimizes a certain distance function $d(\hat{\omega}, \omega)$ or maximizes the posteriori probability $P(\hat{\omega} \mid \omega)$. This problem is referred to as one of *text error correction.*

Different methods for handling the text error correction problem have been proposed in the literature: Statistical methods as in [4]; Dictionary based methods [6]; Hybrid methods (a combination of the previous ones) [5]. All of these methods assumed that each character a_i in the observed word ω is a hard decision output by a classifier. However, the classification of a character depends on the proximity values between a group of prototype characters and the input to be classified [12]. Consequently, it is possible to take into account that information and to provide a fuzzy character $\tilde{a}_i$ $(1 \leq i \leq m)$ for each input, that is, $\tilde{a}_i : \Sigma \rightarrow [0, 1]$.

In [2], a fuzzy method for the CP stage was introduced. The method allows to combine the information of a dictionary together with a sequence of observed fuzzy characters, $\tilde{\omega} = \tilde{a}_1 \tilde{a}_2 \dots \tilde{a}_m$. A simple and flexible implementation of the method was achieved by using Deformed Systems [8]. In [2], the method was only applied to deal with change errors produced by an Isolated Character Classifier (ICC), we denoted these errors as ICC errors. The results show that over a 97% of words with change errors was corrected in an experiment using an ICC of 32% of change error rate, a text of 6396 words, and a dictionary of 1700 words. For that experiment the total recognition rate was 99.10% which improves the results given by other methods [5].

In this paper, a fuzzy method for the text error correction problem is introduced. The method is based on a Deformed System, in particular, a fuzzy automaton is defined to model the possible delete, insert or change errors in the words of texts. The deformed fuzzy automaton is implemented in order to work with sequences of fuzzy characters. The experimental results show a good performance for the three types of errors.

The rest of the paper is organized as follows. Section 2 is devoted to introduce the method. Section 3 provides an efficient algorithm for implementing the method. Finally, the experimental results and the conclusion are shown in Section 4.

2 Deformed Fuzzy Automata

The proposed fuzzy method will begin with the definition of the finite deterministic automaton which accepts (recognizes) the set D. More precisely, we provide an automaton for each word in the dictionary. Given a dictionary $D = \{w_1, \dots, w_r\}$, $\omega_i \in \Sigma^+$; we define the automaton $M(\omega_i)$ in the follow-

Input: $M(\omega_i) = (Q, \Sigma, \delta, \{q_0\}, \{q_n\})$, $Q = \{q_0, \ldots, q_n\}$, n is the length of the chain ω_i
$\tilde{\omega}$ observed string, lenght m ($\tilde{x}_k$ k-th symbol of $\tilde{\omega}$)

Output: C has the value $\mu_{\tilde{L}(MD(\omega_i))}(\tilde{\omega}) = \max_{q',q'' \in Q}(\sigma(q') \star \tilde{\mu}^*(q', q'', \tilde{\omega}) \star \eta(q''))$
where $MD(\omega_i) = (Q, \Sigma, \mu, \sigma, \eta, \tilde{\mu})$ is the deformed fuzzy automaton for $M(\omega_i)$

algorithm initiation
  $\forall i : 1..n$:
    $C(q_i) := 0$;
  $C(q_0) := 1$;
  phase_2
endalgorithm
algorithm computation
  $\forall k : 1..m$:
    phase_1(k);
    phase_2
endalgorithm
algorithm decission
  $C := C(q_n)$
endalgorithm
Note 1: $C(q_{-1}) = 0$.

algorithm phase_1(k)
  $\forall i : 0..n$:
    $C'(q_i) := \max(C(q_i) \star \max_{a \in \Sigma}(\mu_{d_a}), C_1(q_{i-1}, q_i))$;
      where $C_1(q_{i-1}, q_i) = C(q_{i-1}) \star \max_{a \in \Sigma}(\mu_{c_{ax}} \star \mu_{\tilde{x}_k}(a))$,[1]
      and x is that $\delta(q_{i-1}, q_i, x) = 1$
  $\forall i : 0..n$: $C(q_i) := C'(q_i)$
endalgorithm
algorithm phase_2
  $\forall i : 1..n$:
    $C(q_i) := \max(C(q_i), C(q_{i-1}) \star \mu_{i_x})$;
      where x is that $\delta(q_{i-1}, q_i, x) = 1$
endalgorithm

Figure 1: Algorithm for the deformed automaton

ing way. If the word ω_i has length n_i, then $M(\omega_i) = (\{q_0^i, q_1^i, \ldots, q_{n_i}^i\}, \Sigma, \delta^i, q_0^i, \{q_{n_i}^i\})$. The states are ordered from 0 to n_i; where q_0^i and $q_{n_i}^i$ are the start and final state respectively. The transition function is $\delta^i(q_{k-1}^i, q_k^i, a) = 1$, being $a \in \Sigma$ the k-th character in the word ω_i. The rest of the transition function values are null.

The proposed method must handle with erroneous observed words. For that purpose, we define for each $M(\omega_i)$ a fuzzy automaton [7, 9] $MF(\omega_i) = (Q^i, \Sigma, \mu^i, \sigma^i, \eta^i)$ which models the possible insert, delete and change errors for the word ω_i. In the previous definition $Q^i = \{q_0^i, q_1^i, \ldots, q_{n_i}^i\}$ is the set of states; $\sigma^i(q_0^i) = 1$ (0 for the remainder states) is the initial fuzzy configuration, $\eta^i(q_{n_i}^i) = 1$ (0 for the remainder states) is the final fuzzy configuration; and the transition function is defined by the following procedure (the symbol ε denotes the empty string in the expressions):

(1) if $\delta^i(q_{k-1}^i, q_k^i, a) = 1$ then
  $\mu^i(q_{k-1}^i, q_k^i, a) = \mu_{c_{aa}} = 1$ (no error)
  $\mu^i(q_{k-1}^i, q_k^i, \varepsilon) = \mu_{i_a}$ (insert operator)
  $\forall x \in \Sigma$, $x \neq a$: $\mu^i(q_{k-1}^i, q_k^i, x) = \mu_{c_{xa}}$ (change op.)
(2) $\forall q^i \in Q^i$, $\forall x \in \Sigma$: $\mu^i(q^i, q^i, x) = \mu_{d_x}$ (delete op.)

One can note that the different error operators are introduced in order to simulate the inverse effect over the observed word, e.g., a fuzzy transition generated via a delete operator is supplied to model the possible insert error in the observed word. The values $\mu_{i_a}, \mu_{c_{xa}}, \mu_{d_x} \in [0, 1]$ are the membership values associated to the fuzzy transitions, and they can be interpreted as a 'cost' for each transition.

It is possible to expand the transition function μ^i to work with strings from Σ^*, that is, $\mu^{i^*} : Q^i \times Q^i \times \Sigma^* \to [0, 1]$ is defined as (i) $\mu^{i^*}(q', q'', \varepsilon) = \mu^i(q', q'', \varepsilon)$; and (ii) $\mu^{i^*}(q', q'', \omega a) = \max_{q \in Q^i}(\mu^{i^*}(q', q, \omega) \star \mu^i(q, q'', a))$, with $q', q'' \in Q^i$, $a \in \Sigma \cup \{\varepsilon\}$, and $\omega \in \Sigma^*$.

In the above expressions, $\star$ represents the composition functions [13]: $\mu'(x) \star \mu''(x) = \mu'(x) \cdot \mu''(x) / [\lambda + (1 - \lambda) \cdot (\mu'(x) + \mu''(x) - \mu'(x) \cdot \mu''(x))]$.

In our discussion, if $\tilde{L}(MF(\omega_i))$ is the fuzzy language which is accepted by the fuzzy automaton $MF(\omega_i)$, then for a string ω the automaton calculates the following membership value $\mu_{\tilde{L}(MF(\omega_i))}(\omega) = \max_{q', q'' \in Q^i}(\sigma^i(q') \star \mu^{i^*}(q', q'', \omega) \star \eta^i(q''))$.

One of the main motivations to use fuzzy characters as inputs to the CP is due to the fact that an ICC (or a multiple classificator) can supply outputs in the measurement level [12], e.g., each character of the alphabet has a proximity value to the character to be classified. In this work, we will study the same combination as in [2, 3] for handling with the complete type of errors. In the following, we will formulate the deformed fuzzy automaton $MD(\omega_i)$ for the fuzzy automaton $MF(\omega_i)$.

Let $\tilde{x} = \{(a, \mu_{\tilde{x}}(a) \mid a \in \Sigma\}$ be a *fuzzy character*. We assume that it is supplied for the ICC being $\mu_{\tilde{x}}(a)$ the normalized value of proximity of the input x to each character of the alphabet. We denotes $\tilde{\Sigma}$ as the set of the possible fuzzy characters over the universe Σ. Therefore, it is necessary to handle strings of observed fuzzy characters, denoted $\tilde{\omega}$. In the following, we illustrate how to transform the fuzzy automata $MF(\omega_i)$ to accept such strings of fuzzy characters. The deformed fuzzy automaton $MD(\omega_i)$ is obtained from $MF(\omega_i) \equiv (Q^i, \Sigma, \mu^i, \sigma^i, \eta^i)$ via modifying the

Table 1: Word Recognition rates and error correction rates for different composition functions and different word error rates. The experiments have been carried out for the three sets of cost values A, B and C

Dictionary I (1720 words)	Algebraic Product			Einstein Product			Minimum		
(%) Word Error	31.2	44.7	71.7	31.2	44.7	71.7	31.2	44.7	71.7
(%) Recognized Words	A) 95.5	93.2	85.6	A) 95.6	93.2	85.7	A) 78.2	68.3	44.2
	B) 95.5	93.2	85.8	B) 95.5	93.3	85.9	B) 77.8	67.7	43.5
	C) 95.5	93.1	85.7	C) 95.6	93.3	85.9	C) 77.7	67.5	43.4
(%) Corrected Words	A) 85.6	84.8	79.9	A) 85.9	84.8	80.1	A) 30.3	29.3	22.2
	B) 85.8	84.8	80.2	B) 85.7	84.9	80.4	B) 28.9	27.9	21.1
	C) 85.6	84.7	80.0	C) 85.9	85.0	80.4	C) 28.8	27.5	20.9
(%) Corrected ICC Errors	A) 94.4	94.1	91.8	A) 94.8	94.3	92.0	A) 67.6	64.3	46.5
	B) 93.2	92.8	90.5	B) 93.4	92.8	90.7	B) 60.4	56.5	40.8
	C) 93.7	93.2	91.1	C) 93.9	93.4	91.4	C) 58.5	54.2	39.0
(%) Corrected Insert Errors	A) 89.4	89.0	85.9	A) 89.9	89.4	86.2	A) 00.1	00.2	00.5
	B) 90.5	89.8	87.0	B) 90.7	90.3	87.6	B) 00.1	00.2	00.5
	C) 88.6	87.8	84.9	C) 89.4	88.5	85.5	C) 00.1	00.1	00.4
(%) Corrected Delete Errors	A) 75.2	74.7	71.3	A) 75.6	74.6	71.5	A) 41.2	37.9	24.7
	B) 77.0	76.2	73.7	B) 76.5	76.2	73.8	B) 48.0	44.8	29.7
	C) 77.5	77.0	74.6	C) 77.6	77.3	74.8	C) 50.2	46.3	31.5
(%) Corrected Change Errors	A) 84.5	84.0	81.1	A) 84.5	83.7	81.0	A) 13.3	12.6	09.1
	B) 83.7	83.2	80.4	B) 83.3	83.1	80.1	B) 10.5	09.8	06.9
	C) 84.1	83.5	80.8	C) 84.0	83.6	80.8	C) 10.1	09.3	06.6

transition function. The new transition function, denoted $\tilde{\mu}^i$, $\tilde{\mu}^i : Q^i \times Q^i \times (\tilde{\Sigma} \cup \{\varepsilon\}) \rightarrow [0,1]$ is defined as follows:

(1) $\tilde{\mu}^i(q', q'', \tilde{a}) = \max_{x \in \Sigma}(\mu^i(q', q'', x) \star \mu_{\tilde{a}}(x))$, with $q', q'' \in Q^i$, and $\tilde{a} \in \tilde{\Sigma}$;
(2) $\tilde{\mu}^i(q', q'', \varepsilon) = \mu^i(q', q'', \varepsilon)$, with $q', q'' \in Q^i$.

It is possible the extension of $\tilde{\mu}^i$ to $\tilde{\mu}^{i^*} : Q^i \times Q^i \times \tilde{\Sigma}^* \rightarrow [0,1]$, in the same way as in the fuzzy automaton.

If $\tilde{L}(MD(\omega_i))$ is the fuzzy language which is accepted by the deformed fuzzy automaton $MD(\omega_i)$, then for an observed fuzzy string $\tilde{\omega}$ the machine calculates the membership value $\mu_{\tilde{L}(MD(\omega_i))}(\tilde{\omega}) = \max_{q', q'' \in Q^i}(\sigma^i(q') \star \tilde{\mu}^{i^*}(q', q'', \tilde{\omega}) \star \eta^i(q''))$.

Therefore, given an observed string of fuzzy characters $\tilde{\omega} \in \tilde{\Sigma}^*$, each $MD(\omega_i)$ calculates $MD(\omega_i, \tilde{\omega}) = \mu_{\tilde{L}(MD(\omega_i))}(\tilde{\omega})$, which can be interpreted as the 'cost' to transform $\tilde{\omega}$ in ω_i. Finally, $\tilde{\omega}$ is classified as a dictionary word $\hat{\omega} \in D$, such that $MD(\hat{\omega}, \tilde{\omega}) \geq MD(\omega_i, \tilde{\omega})$, $\forall \omega_i \in D$.

3 The Algorithm

In practice, it is possible to development an algorithm that simulates the process to be carried on by the deformed fuzzy automaton. We does not explicitly make the $MD(\omega_i)$ automaton for each ω_i in the dictionary; the algorithm operates on the original automata $M(\omega_i)$.

The algorithm for performing the process has to compute the final membership value associated to every one of the strings that can be obtain from the observed fuzzy word $\tilde{\omega}$. These strings are obtained by applying repeatedly the insert, delete and change operations. Finally, we choose as the estimated string, the string with the minimum cost (maximum membership). The execution time needed for this process is very high. Thus, for example, for the change operation it would be necessary to process t^m strings, where t is the size of Σ and m is the length of the observed fuzzy word. However, in order to decrease the execution time, we have developed a dynamic programming algorithm (similar to that proposed by Viterbi [10, 11] and also by Bouloutas [1]). This algorithm computes for every input fuzzy character, the fuzzy transition needed for reaching every state of the automaton. Figure 1 shows this algorithm. Given an automaton $M(\omega_i)$, the complexity of the *phase_1* is $\mathcal{O}(t^2 \cdot n)$, where t is the size of Σ and n is the length of the string ω_i. In the other hand, the *phase_2* has a complexity $\mathcal{O}(n)$. The complete execution of the algorithm *computation* for an observed string of length m is $\mathcal{O}(m \cdot t^2 \cdot n)$.

4 Experimental Results

Basically the experiments consist in processing handprinted texts by the text recognition system and comparing the results before and after the deformed fuzzy automaton stage; thus, we can analyze how efficient is this automaton to reduce the error rate.

First, we have tested the recognition system with a large collection of values for the deformed automaton transitions. From these previous experiments we con-

Table 2: Word Recognition rates and error correction rates for different dictionary sizes

	Dictionary II (3158 words)			Dictionary III (6838 words)		
(%) Word Error	31.2			31.2		
(%) Corrected Words	A) 82.3	B) 82.3	C) 82.4	A) 78.4	B) 78.6	C) 78.6
(%) Corrected ICC Errors	A) 93.4	B) 91.4	C) 92.3	A) 91.2	B) 89.5	C) 90.2
(%) Corrected Insert Errors	A) 87.1	B) 87.8	C) 86.0	A) 82.4	B) 83.6	C) 81.4
(%) Corrected Delete Errors	A) 69.0	B) 71.1	C) 71.8	A) 65.2	B) 67.3	C) 68.5
(%) Corrected Change Errors	A) 81.3	B) 80.6	C) 81.1	A) 76.8	B) 75.9	C) 76.1

clude that these values may verify certain relations among them to obtain good performances. We have chosen three set of values that verify such relations for the experiments: $\forall x, y \in \Sigma$ A) μ_{i_x}= 0.001 $\mu_{c_{xy}}$= 0.0005 μ_{d_x}=0.0001; B) μ_{i_x}= 0.005 $\mu_{c_{xy}}$= 0.001 μ_{d_x}= 0.0005; C) μ_{i_x}= 0.01 $\mu_{c_{xy}}$= 0.001 μ_{d_x}= 0.0001;

Table 1 shows the results obtained in the experiments for different composition functions (Algebraic product $\lambda = 1$, Einstein product $\lambda = 2$, and minimum) and for different edition error rates introduced by both the segmentation process and the ICC (% word error in the table). In these experiments we have used a 1000-words text written by many different authors and a 1720-words dictionary. In the table 1, we can see the rate of recognized words and also the rates of corrected errors for every error type. The results are very similar for the *algebraic product* and for the *Einstein product.* These values are very high; thus, even for a word error of 71.5%, the deformed fuzzy automaton is able to correct about an 80% of the erroneous words. However, the results obtained with the *minimum* function are very poor. Anyway, this result was expected because the minimum is a very drastic operator. In the other hand we can see how different correction rates are obtained depending on the kind of errors. Thus, best results are obtained in the case of ICC errors. This can be explained because the automaton, when it corrects a ICC error, can uses the measurement level that the ICC provides for an input letter. Finally, we can also see how the results are very similar for the three sets of cost values selected.

Table 2 shows the results obtained with the same 1000-words text but with different dictionary sizes: a 3158-words dictionary and a 6838-words dictionary. We can see in the table how the results are now lower than those in table 1. However, they are still high. Thus, with the largest dictionary the automaton corrects more than the three quarter parts of the erroneous words. In conclusion, the experimental results show good performance in correcting the three types of errors.

References

[1] A. Bouloutas, G.W. Hart, M. Schwartz; Two Extensions of the Viterbi Algorithm; *IEEE Tran. on Information Theory*, 37(2) pp. 430-436; 1991.

[2] J. Echanobe, J.R. González de Mendívil, J.R. Garitagoitia, C. Alastruey; Deformed Systems for Contextual Postproccesing; *Fuzzy Sets and Systems* 96, pp. 335-341; 1998.

[3] J. Echanobe. Tesis Doctoral. Universidad Pública de Navarra. 1998.

[4] A. Goshtasby, R.W. Ehrich; Contextual Word Recognition Using Probabilistic Relaxation Labeling; *Pattern Recognition* 21(5), pp. 455-462; 1988.

[5] J.J. Hull, S.N. Srihari, R. Choudhari; An Integrated Algorithm for Text Recognition: Comparison with a Cascaded Algorithm; *IEEE Trans. Pattern Analysis Mach. Intell.* 5(4), pp. 384-395; 1983.

[6] R.L. Kashyap, B. J. Oommen; Spelling Correction Using Probabilistic Methods; *Pattern Recognition Letters* 2(4), pp. 147-154; 1984.

[7] M. Mizumoto, J. Toyoda, K. Tanaka; Fuzzy Languages; *Systems Computers Control*; pp. 36-43; 1970.

[8] C.V. Negoita, D.A. Ralescu; *Application of Fuzzy Sets to System Analysis*; Birkaeuser, Basilea; 1975.

[9] M.G. Thomason; Finite fuzzy automata, regular fuzzy languages and pattern recognition; *Pattern Recognition* 5; pp. 383-390; 1973.

[10] A.J. Viterbi; Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm; *IEEE Tran. on Information Theory* 13(2); pp. 260-269; 1967.

[11] A.J. Viterbi; Convolutional Codes and Their Performance in Communication Systems; *IEEE Tran. on Communications Technology* 19(5); pp. 751-772; 1971.

[12] L. Xu, A. Krzyzak, C.Y. Suen; Methods for Combining Multiple Classifiers and Their Applications to Handwritting Recognition; *IEEE Trans. Sys. Man and Cyber.* 22(3); pp. 418-435; 1992.

[13] H.J. Zimmermann; *Fuzzy Set Theory and its Applications*; Kluwer Academic Pub., Boston, MA; 1990.

FUZZY COLOUR DISTANCE APPLIED TO REGION GROWING IN IMAGE PROCESSING

R. Soria
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid

A. Sanchis
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid
masm@inf.uc3m.es

J. M. Molina
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid molina@ia.uc3m.es

Abstract

In this work, a new approach to region growing problem is presented. The region growing procedure is modified in order to apply a fuzzy knowledge based system to evaluate the distance between a pixel and the region. This fuzzy colour distance is based on a set of rules that evaluate the three distance components (R, G, B) jointly. The results presented show a improvement with the traditional crisp distance.
Keywords: Image Processing, Fuzzy Distance, Knowledge Based Systems.

1 INTRODUCTION

The objects tracking in images sequences obtained through a camera, located in a fixed position is a problem that could be approached with a classic tracking filter. In order to accomplish the tracking process in image, the definition of the interest zones for each acquired image is required [3]. In this work, a region growing procedure is use to obtain a set of interest zones that are moving at time and therefore they appear in different positions in each one of the images. The extraction of the information of the interest zones in movement can be obtained through the segmentation from each acquired image [6]. This procedure has a high computational cost, since each acquired image should be processed for finally only use concrete segments in the tracking process. In this case a high amount of information is required and probably any machine is able to accomplish the processing of all images in an acceptable time. However, the quantity of information of the sequence is reduced if only a subset of global image is considered, the neighbourhood in an environment of zones. Then, once the zones that are found in movement are determined, the information of the position of the zone in an image can be used in the following image to obtain the region on which to accomplish the segmentation, reducing the calculation space and therefore the execution time.

Each zone remains determined by a set of parameters, on those which is produced a filtered and a prediction that allows to locate the neighbourhood window in the following instant. The used filter is a filter α - β. The proposed scheme, in general, can be observed in Figure 1.

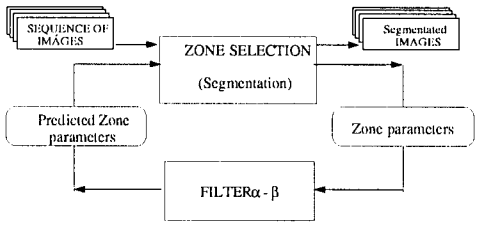


Figure 1: Scheme of segmentation and prediction system

The segmentation scheme is based on region growing [1]. The algorithm searches inside a image regions that have the same type of parameters. In this way, a zone is defined by the set of pixels that has the same characteristic.

In traditional algorithms [2], the equality restriction (zero distance) is very strict, and they use likelihood intervals between the pixels to define the belonging of a given zone. In this work, the problem of calculating the distance between a pixel and the value of the zone is approached through a fuzzy system [4]. These systems will be able of considering the distance between the value of a pixel and the one of the zone through a knowledge base that reflect a diversity of options greater than using the classic method of the difference in absolute value of the colours.

2 DISTANCE AND SEGMENTATION

To accomplish the region growing, the method departed from an complete image on which certain windows are predicted. The windows are located in the zones that are needed to extract from the image for tracking process.

The existence of overlapping between windows is not relevant for the problem, since zones are considered independent. The independence is reached by the following method: on the original image is opened a window and, a copy of the values within the same is accomplished in an array. The original is not modified in any case, therefore if other window is overlapped with the previous, the copy has the original values to accomplish the segmentation.

2.1 PIXELS DISTANCE

The definition of the characteristics of a pixel is based on the colour. The colour of a pixel is defined through a vector of three components (R, G, B) [2] and the distance used between the pixels will be the distance between each one of the coordinates, thus the "likelihood" between two pixels, A (R_A, G_A, B_A) and B (R_B, R_B, B_B), is calculated in this way:

$$d_R = R_A - R_B$$
$$d_G = G_A - G_B$$
$$d_B = B_A - B_B$$

Two pixels are equals, or a pixel and the zone characteristics, when any colour distance is over certain threshold. Typically the threshold is located in the interval from 10 to 20 [1].

2.2 SEGMENTATION INSIDE A WINDOW

Once a piece of image where the zone is found, the segmentation process is accomplished. The method chooses pixels of the image and compares the parameters of the pixel with ones of the zone. If the comparison result gives that the pixel is similar to the zone, the pixel is included in the zone and the new parameters of the zone are calculated making the average of all the colour values of the zone pixels [2]:

- A zone is defined by a set of $pixels_{i,j}$ with colour values (R_{ij}, G_{ij}, B_{ij}) }
- The number of pixels that belong to the zone is n
- Then the zone parameters are: $R_{zone} = (\Sigma R_{ij})/ n$, $G_{zone} = (\Sigma G_{ij})/ n$, $B_{zone} = (\Sigma B_{ij})/ n$.

In the first step of algorithm, the zone is composed of a pixel, therefore the zone parameters are the same that the pixel ones: the pixel colour RGB. In the next step new pixels are included and the method must re-define the value of parameters with new pixel colours. The iterative process finishes when surrounding the zone do not exist pixels that could belong to such zone [2]. In this procedure two types of problems should be outlined, on one hand the growing in zones that suffer a demoted soft from colour and additionally the concave zones.

In the first case, if a zone is growing following colours that are removed gently from the first zone value, the zone parameters could be evolving with the new points towards a final value that is far from the real one. The real value is obtained using only the pixels that reality belong to the zone (Figure 2). To solve this problem, the growing must follow a path in spiral so that the withdrawal by the central zone will be gradual and the change in a peripheral part of the zone does not produce a change of the same (Figure 3).

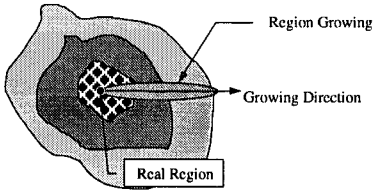


Figure 2: Straight Line Growing in a Image with Degraded Colours

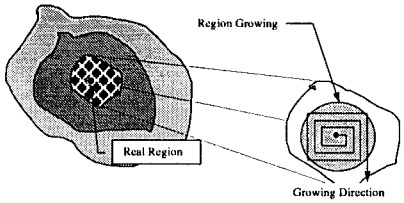


Figure 3: Spiral Growing in a Image with Degraded Colours

In concave zones, some pixels could remain outside the zone (Figure 4). In order to avoid this problem a new region growing method is developed based on spiral growing but always considering the shape of the region that grows.

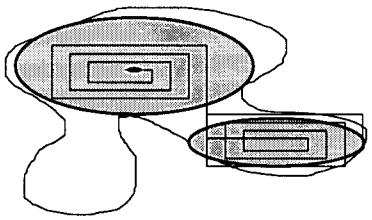


Figure 4: Concave Zone Problems in Region Growing

3 DISTANCE EVALUATION USING A FUZZY SYSTEM

The proposed fuzzy system has as input variables the distance in each colour component of the pixel and the zone. Depending on the value of these distance (Small, Medium or High) and applying the rules of system, a linguistic variable, <Distance> is obtained. This variable is defined by three membership functions (Small, Medium or High). The defuzzification process determines a value that defined the final distance [7]. If this distance is less than certain value, the pixel would be included in the region. The method to calculate the new parameters of the region, including the new pixel, would be the same that in the traditional algorithm.

3.1 INPUT AND OUTPUT VARIABLES

The fuzzy sets that define each input would be identical for each variable (d_R, d_G, d_B) and they would be defined by three levels (**S**mall, **M**edium or **H**igh) that could be

observed in Figure 5. Each input variable represents the difference in absolute value between a pixel and the centre of the region, this difference is the value that is fuzzified to apply the knowledge base, the implication used is the Mamdani one[5], and to obtain the defuzzified distance. The fuzzy sets of output would be defined three levels (**S**mall, **M**edium or **H**igh) that could be observed in Figure 6.

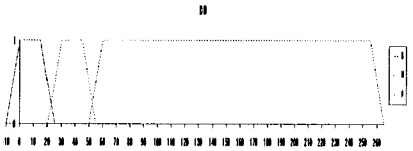


Figure 5: Input Membership Functions

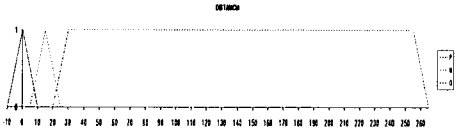


Figure 6: Output Membership Functions

3.2 RULE BASE

The system rules are shown in Tables 1, 2 and 3.

		DG		
		S		
DR	S	S	S	H
	M	S	M	H
	H	H	H	H
		S	M	H
		DB		

Table 1: Distance Rules for Small Green.

		DG		
		M		
DR	S	S	S	H
	M	M	M	H
	H	H	H	H
		S	M	H
		DB		

Table 2: Distance Rules for Medium Green.

		DG		
		H		
DR	S	H	H	H
	M	H	H	H
	H	H	H	H
		S	M	H
		DB		

Table 3: Distance Rules for High Green.

4 RESULTS AND CONCLUSIONS

The experiments carried out has been proven the segmentation algorithm on different images changing the threshold from which is considered that pixel belongs to a region or not. With the traditional distance, considering the absolute value, thresholds of 10, 15 and 20 has been considered. With the fuzzy calculation of the distance this threshold must be function of the centroids of the fuzzy sets that define the labels of the output linguistic variable. In this case the thresholds of experiments have been: 0, 2.5, 5, 7.5, 10, 12.5 and 15. In Figure 7, 8 and 9 are shown the results for one of the images.

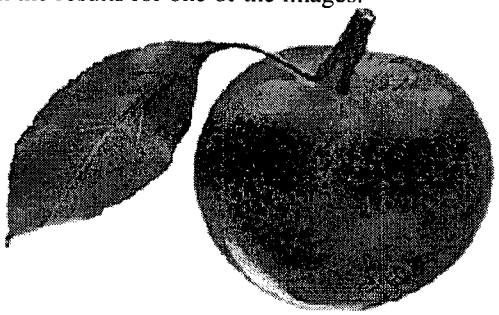


Figure 7: Original Image

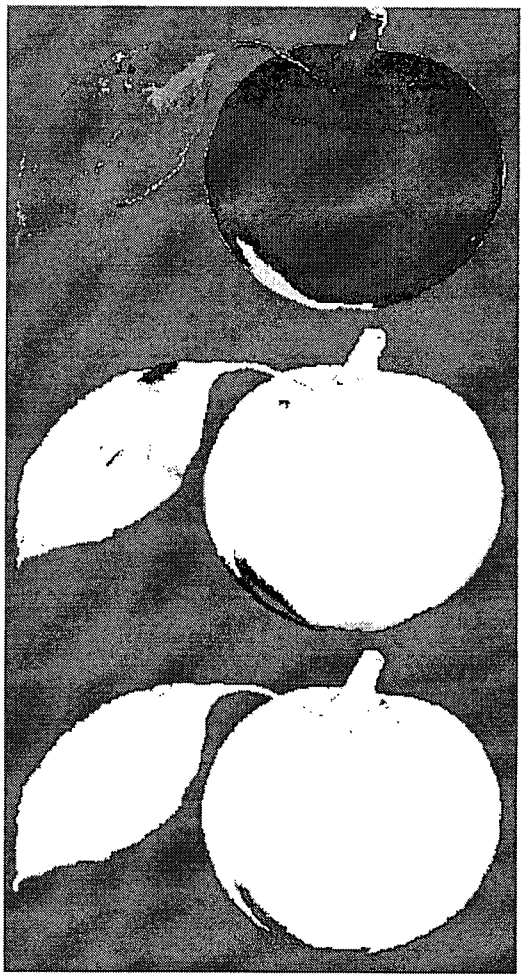
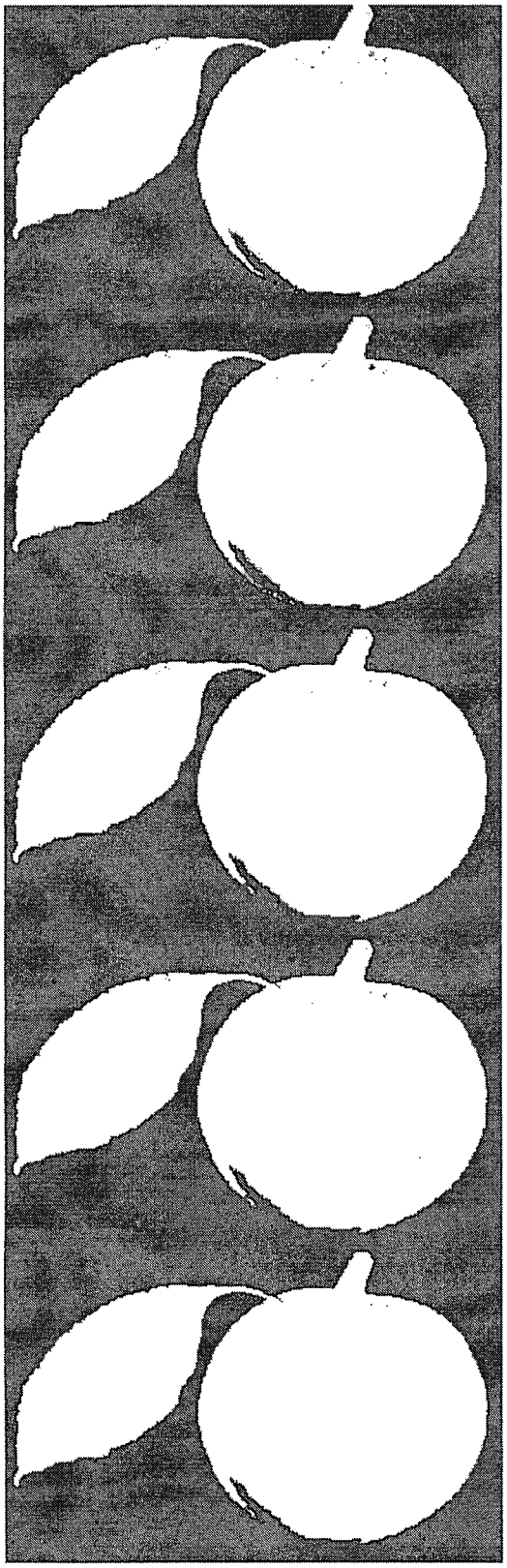


Figure 8: Results for traditional distance with thresholds of 10, 15 y 20 respectively



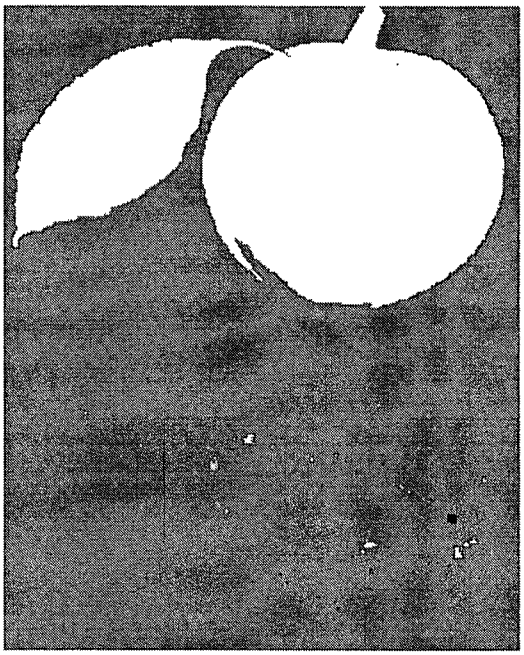


Figure 9: Results for fuzzy distance with thresholds of 0, 2.5, 5, 7.5, 10, 12.5 y 15 respectively

In Figure 8 and 9 the results show that the fuzzy distance produce less zones than traditional one. This result is relevant in order to avoid irrelevant zones with few pixels that produce something like noise, see Figure 9. Following the evolution of image in Figure 9 and 10, the results show that a threshold is the most important parameter of the algorithm, because if the threshold is high no regions grow in the image, and if the threshold is low many regions produce a blurry image.

References

[1] A. K. Jain, *Fundamentals of Digital Image Processing*, Ed. Prentice-Hall, 1989.

[2] Ballard D.H. and Brown C.M., *Computer Vision*, Ed. Prentice-Hall, 1982.

[3] Bertold Klaus Paul Horn, *Robot Vision*, MIT Press 1986.

[4] L. A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", Trans. on SMC, Jan. 1973.

[5] Mamdani H., "Aplication of Fuzzy Logic to Aproximate Reasoning Using Linguistic Synthesis", IEEE Trans. on Computers, Nov/Dic. 1988.

[6] W.K. Pratt, *Digital Image Processing*, john-Wiley & Sons 1991.

[7] Zimmermann H.J., *Fuzzy Sets, Uncertainty and Information*, Kluwer 1985.

Soft Computing as a Methodology for Color Processing

Lars Hildebrand
University of Dortmund
Department of Computer Science
Chair 1, Otto-Hahn-Str. 16
44227 Dortmund 50, Germany
e-mail: hildebrand@ls1.cs.uni-dortmund.de

Madjid Fathi
University of Dortmund
Department of Computer Science
Chair 1, Otto-Hahn-Str. 16
44227 Dortmund 50, Germany
e-mail: fathi@ls1.cs.uni-dortmund.de

Summary

High successful applications of soft computing and fuzzy logic can be found in engineering disciplines. Current research is done for computational subject, e. g. computing with words, or computing with colors, which is one of the new ideas for the processing of color information. The main fact of this approach is the adequate representation of color and its linguistic description. This approach can be applied to different engineering fields. This work continues the idea of linguistic color processing, that we presented in [8] for the first time.

Keywords: Color, Color Models, Fuzzy Logic

1 INTRODUCTION

Color is a human perception, created by light of specific wavelengths that falls on different cones in the eye's retina. There are three types of cones that approximately correspond to red, green and blue lights and one additional type that corresponds to the brightness. As a result, color is created in our brain, when we see a distribution of light of different wavelengths. This individual perception and also the individual naming of colors makes it difficult to use knowledge that is based on colors in technical applications.

A adequate representation of colors and the use of fuzzy logical for a linguistic and vague description is described here to overcome the stated problems.

2 COLOR MODELS

Many color representations are based on technical demands. Examples are the red, green, blue division of colors for televisions or the cyan, magenta, yellow division for the printing media. These two representations reflect the additive and subtractive mixing of few base colors to obtain a large set of displayable or printable colors. Technical representations are suitable for displaying colors, but fail if a deeper understanding of color is needed [10].

One field of applications is the knowledge based reasoning, where knowledge of human experts is acquired to set up systems for decision support. Human description of colors is not based on the additive or subtractive mixing of base colors [5, 7]. It is more oriented by characteristics like hue, brightness or lightness. Apart from technical color models another class of color models exist, that fulfills these human demands.

Modeling colors in this way is more like an artist view of mixing colors: Starting from a pure color (hue), white pigments are used to lighten the color (decrease of saturation), and black pigments are used to darken the color (decrease of lightness). One of the earliest of these color models was the one originated by A. H. Munsell in 1915 [4, 6].

The interesting fact in this color model is that Munsell arranged the hues in a 360° polar coordinate system. The same arrangement of colors is used in the HSI-color model, which is similar to Munsell's color model.

The HSI-model is well suited for the linguistic processing of color, because colors that are similar for humans are grouped together and there is a clear distinction of colors and grays. The 3-dimensional model can be easily reduced to a 2-dimensional one, by simply dropping the intensity coordinate when the pure and light colors are important, or by dropping the saturation coordinate when pure and dark colors are important. These operation result in the HS- and HI-sub models, that are used throughout approach [8].

2.1 COLOR MODEL CONVERSION

A conversion from the RGB-color model to the HSI-color model can be easily done by using the following algorithm [3, 11]. At first some constant values have to be defined: The unit vector $\vec{I}_u$, the red vector $\vec{R}$, and the green unit vector $\vec{G}_u$ are defined:

$$\vec{I}_u = \left\| \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\|, \quad \vec{R} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \tag{1}$$

$$\vec{G}_u = \left\| \vec{I}_u \times \vec{R} \right\| = \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} \tag{2}$$

$$\vec{R}_u = \left\| \vec{R} \right\| = \begin{pmatrix} \frac{\sqrt{6}}{3} \\ -\frac{\sqrt{6}}{6} \\ -\frac{\sqrt{6}}{6} \end{pmatrix} \tag{3}$$

A color $\vec{c}$ that is given as

$$\vec{c} = \begin{pmatrix} R \\ G \\ B \end{pmatrix}$$

where R, G, B denote the red, green, and blue part of the color can then be converted into hue, saturation and intensity using

$$I(\vec{c}) = \frac{1}{\sqrt{3}} \vec{c} \cdot \vec{I}_u \tag{4}$$

$$S(\vec{c}) = 1 - \frac{\min(\vec{c})}{I(\vec{c})} \tag{5}$$

$$H(\vec{c}) = \operatorname{atant}\left(\vec{c} \cdot \vec{G}_u, \vec{c} \cdot \vec{R}_u\right) \tag{6}$$

The operations $\|\vec{v}\|$, $\vec{v}_1 \cdot \vec{v}_2$, and $\vec{v}_1 \times \vec{v}_2$ stand for normalization, vector dot, and vector cross product.

3 SOFT COMPUTING APPROACH

The main idea of fuzzy logic and fuzzy logic based based color processing can be found in [1, 8]. Like other fuzzy systems, the automated generation of the necessary fuzzy sets is not simple. An automated generation, based on the combination of evolution strategies and fuzzy logic [2] can be used, if a sufficient number of supporting points can be found.

To calculate these supporting points a statistical analysis of the image colors and a representation of the color distribution as a histogram is needed. Examples are shown in figure 1 and 2.

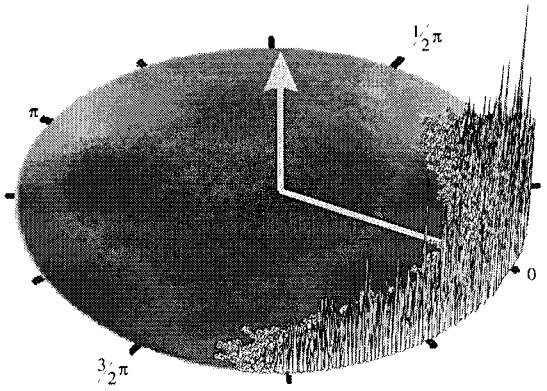


Figure 1: Color histogram 1

The shown histograms are results of the processing of some samples. These histograms are frequency distributions of the used colors in those images. The histograms can directly be used to calculate the membership of a given color. The only calculation needed is the normalization of the frequencies to the interval [0, 1].

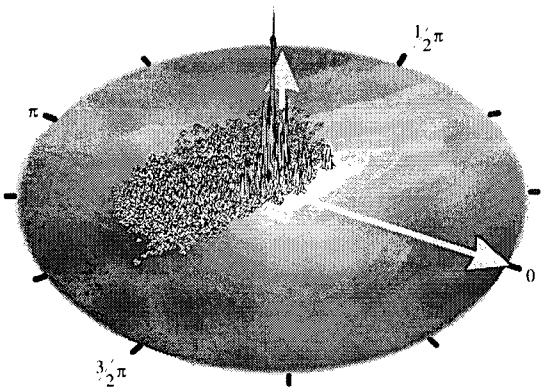


Figure 2: Color histogram 2

Another approach is the automated generation of fuzzy sets, that corresponds to the given frequency distribution. These fuzzy sets should be of the shape described in [8]. An example for the base color red, green and blue, defined this way shows figure 3.

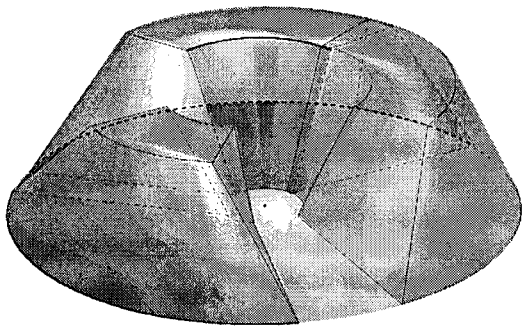


Figure 3: Definition of red, green, and blue

One way to generate them is the extension of our algorithm described in [2] from 1-dimensional fuzzy sets to 2-dimensional ones. This generation of 2-dimensional fuzzy sets may not be confused with the generation of 2-dimensional fuzzy rulebases. The first method, that is needed here, is the generation of fuzzy sets defined over a 2-dimensional universe. The latter one is the generation of a fuzzy rulebase system with 2 input variables.

4 APPLICATION EXAMPLE

Resistance spot welding is a welding process that uses the inherent resistant of metal workpieces to join two sheets of metal by the flow of electrical current. This current is passed trough the metal sheets and generates the welding heat. The same electrodes that supply the current, apply simultaneously a force and the two metal parts are joined.

Most of the quality testing systems have to destroy the welding joint in order to obtain quality measures like longitudinal and transverse tensile strength, bend strength, or hardness. The quality criteria of a single welding spot can be divided into two major ranges: physical and optical criteria. Physical criteria, like tensile or bend strength describe the physical properties of a welding joint. Most important quality criteria can be determined using color information of the welding spot. Some preprocessing using standard methods from the field of image processing may be necessary [9].

Using the idea of fuzzy color description only few colors (blue, red, and gray) have to defined. The use of modifiers allows to automatically generate corresponding color definitions like light red, dark red, light blue, and dark blue. Transforming the experts knowledge about the colors into a fuzzy rulebase leads to the following set of rules:

IF COLOR_OF_INNER_SPOT IS BLUE
THEN QUALITY IS GOOD

IF COLOR_OF_OUTER_SPOT IS LIGHT_BLUE
THEN QUALITY IS GOOD

IF COLOR_OF_IMPACT_ZONE IS DARK_RED
THEN QUALITY IS GOOD

IF COLOR_OF_INNER_SPOT IS LIGHT_RED
THEN QUALITY IS POOR

IF COLOR_OF_OUTER_SPOT IS BLUE
THEN QUALITY IS POOR

IF COLOR_OF_OUTER_SPOT IS RED
THEN QUALITY IS POOR

IF COLOR_OF_IMPACT_ZONE IS LIGHT_RED
THEN QUALITY IS POOR

This set of rules can be evaluated using standard fuzzy techniques in combination with the described methods to calculate the membership values of colors. Any fuzzy system that allows to add new functions for the calculation of membership values can be extended to benefit from fuzzy color processing.

5 SUMMARY AND OUTLOOK

Soft computing for color processing is a powerful approach. Problems with the generation of the corresponding fuzzy sets can be solved using the two mentioned methods. Only few rules are needed to solve complex problems. In our example, the evaluation of a total of seven rules is enough to distinguish between welding points with a good and those with a bad quality. The computational demand is low, so cheap microcontrollers can be used to solve this task. The ideas of linguistic modifiers can be integrated, to generate other color definitions like *light_red*, *dark_red* or *more_or_less_red* automatically from the definition of the color *red*. This will reduce the development time for new applications.

References

[1] D. Dubois, H. Prade, Y. Y. Yager (1993) Fuzzy sets for intelligent systems. Morgan Kaufmann Publishers, San Mateo.

[2] M. Fathi, L. Hildebrand (1997) Model-free optimization of fuzzy rulebased systems using evolution strategies. In*: IEEE Transactions on System, Man, and Cybernetics, Vol. 27, No. 2.*

[3] J. Foley, et al. (1990) Computer graphics: Principles and practice. 2nd Edition, Addison-Wesley, Reading.

[4] Macbeth (1979) The Munsell book of color. Macbeth, a division of Kollmorgen Corp., Baltimore.

[5] M. V. McLean (1965) Brightness contrast, color contrast, and legibility. *Human Factors, No. 7.*

[6] A. H. Munsell (1946) A color notation. Munsell Color Company, Baltimore.

[7] G. Murch (1986) Color displays and color science. In: I. Durret, H. John (editors) *Color and the computer*, Academic Press, San Diego, California.

[8] B. Reusch, M. Fathi, L. Hildebrand (1998) Fuzzy Color Processing for Quality Improvement. In: *Proceedings International Forum on Multimedia and Image Processing*, World Automation Congress, Anchorage.

[9] A. Rosenfeld, A. C. Kak (1982) Digital picture processing. Vol. 1 and 2, 2nd Edition, Academic Press, San Diego.

[10] L. G. Thorell (1983) Introduction to color vision. In: *Proceedings of the society of photo-optical instrumentation engineers.*

[11] A. H. Watt (1989) Fundamentals of three-dimensional computer graphics. Addison-Wesley, Wokingham.

Criterios de Parada Difusos en el Problema de la Mochila

J.L. Verdegay
Depto. de CC. de la Computación e I.A.
Universidad de Granada
E–18071 Granada
verdegay@goliat.ugr.es

E.R. Vergara
Depto. de Matemáticas
Universidad Nacional de Trujillo
Trujilllo-Perú
edmundo@goliat.ugr.es

Resumen

En este artículo se introducen criterios de parada difusos en los algoritmos del Problema de la Mochila, y como consecuencia se obtiene dos nuevos algoritmos, cuyo funcionamiento se ilustra con algunos ejemplos.

Palabras Clave: Algoritmos, Problema de la Mochila, Reglas difusas.

1 Introducción

Como es de sobra conocido, las metodologías asociadas a los conjuntos difusos se han apoyado prácticamente siempre en las que previamente existían (clásicas), siendo poco frecuente el que los modelos convencionales se traten de resolver a partir de métodos originales del campo difuso. Sin embargo esto no se ha dado con los Sistemas Basados en Reglas, que si han aportado una metodología propia proveniente de los conjuntos y los sistemas difusos. Dentro del campo de la Inteligencia Artificial, y también en otros campos, una de las áreas más provechosas, en diferentes sentidos, es la de los modelos de programación matemática, y dentro de estos, uno de los problemas más relevantes, tanto por sus aplicaciones como por ser un autentico banco de pruebas metodológicas, es el Problema de la Mochila (PM). Desde este doble punto de vista, el ámbito en el que se desarrolla este trabajo es el del interfaz entre los Sistemas Basados en Reglas y el PM.

El PM consiste en llenar una mochila de forma que se optimice el valor de los objetos transportados, respetando una limitación de capacidad. Este problema es NP-completo [3], es decir, que no se conoce un algoritmo polinomial, que obtenga su solución óptima. En estas circunstancias, el uso de los algoritmos aproximados, o heurísticas, es muy interesante porque permiten obtener soluciones que si bien no son óptimas, si están cercanas a ellas. De cara a su solución, también es una alternativa, tanto para problemas NP-completos como para el PM en particular, considerar criterios de parada difusos (basados en reglas difusas) en los algoritmos exactos conocidos.

Clásicamente los criterios de parada fijan las condiciones de finalización del procedimiento iterativo de un algoritmo, estableciéndose dichos criterios a partir de las características teóricas del problema, del tipo de solución que se busca y del tipo del algoritmo que se utiliza, que en definitiva determina un conjunto de referencia, y se detiene cuando se verifica el criterio de parada.

La flexibilización de los algoritmos exactos con la introducción de criterios de parada basados en reglas difusas, supone considerar que el conjunto de referencia es un conjunto difuso, y los criterios de parada difusos se fijan en función del grado de pertenencia de los elementos. Estos criterios de parada difusos fueron introducidos por primera vez en el ámbito de la programación matemática en [5], y luego desarrollados en [1] para el caso particular del algoritmo simplex. En [6] se han generalizado los resultados de [1] y [5] usando funciones de pertenencia no lineales, y se han estudiado criterios de parada difusos para el algoritmo de Karmarkar y sus extensiones.

Se sabe que en los algoritmos exactos del PM, los criterios de parada varían de un algoritmo a otro, pero al introducir estos criterios difusos uniformizamos la diversidad de los criterios de parada convencionales, en la siguiente sección recordamos algunos algoritmos exactos y un algoritmo para la obtención de cotas superiores del PM. En la sección 3 presentamos los criterios de parada difusos, y en la sección 4, los resultados numéricos y las conclusiones.

2 Algoritmos para el Problema de la Mochila

Formalmente planteado, el PM 0-1 consiste:

$$Max\left\{\Sigma_{j\in J} p_j x_j : \Sigma_{j\in J} w_j x_j \leq c;\ x_j \in \{0,1\}, j\in J\right\} \qquad (1)$$

siendo J={1,...,n} el conjunto de índices que referencia los objetos o ítems para rellenar la mochila, c la capacidad de la mochila, w_j, y p_j $j\in$ J el *peso* y *valor* respectivamente del j-ésimo ítem, y x_j $j\in$J, tomando valor 1 si el ítem j es seleccionado para introducirlo en la mochila y 0 en otro caso. Convenimos que los ítems satisfacen:

$$\frac{p_1}{w_1} \geq \frac{p_2}{w_2} \geq \ldots \geq \frac{p_n}{w_n}. \qquad (2)$$

y en lo que sigue, notaremos por z(PM) a la función objetivo del PM.

2.1 Cota superior de Dantzig:

Una cota superior U de (1) es tal que z(PM)≤U. Existen varios algoritmos para determinar diferentes cotas superiores [3]. Por su simpleza aquí utilizaremos la cota superior de Dantzig, obtenida a partir de la solución de la relajación continua (C(PM)) que es el mismo problema (1) con la condición $x_j \in [0,1]$. Conociendo el ítem crítico

$$s = \min\left\{ j: \sum_{i=1}^{j} w_i > c \right\} \qquad (3)$$

la cota de Dantzig es el mayor entero del valor óptimo z(C(PM)) de C(PM), es decir,

$$U = \lfloor z(C(PM)) \rfloor = \sum_{j=1}^{s-1} p_j + \lfloor \bar{c}\, p_s / w_s \rfloor \qquad (4)$$

$$\bar{c} = c - \sum_{j=1}^{s-1} w_j \qquad (5)$$

donde $\lfloor x \rfloor$ nota el mayor entero del valor x.

2.2 Algoritmos exactos

Como comentamos, existen varios algoritmos para resolver el PM. Una muy buena recopilación de dichos algoritmos puede encontrarse en [3]. En cualquier caso aparte de los métodos "greedy" conocidos, los enfoques más relevantes son los de ramificación y acotación (Branch and Bound), y el de la Programación Dinámica. Por eso en este trabajo hemos elegido el algoritmo de Horowitz-Sahni, del tipo de ramificación y acotación, y un algoritmo basado en Programación Dinámica para ilustrar la propuesta de uso de los criterios de parada difusos.

2.2.1 El algoritmo de Horowitz-Sahni

El algoritmo de Horowitz y Sahni (HS) [2], que se destaca por su eficiencia, elegancia y simplicidad de implementación, sigue la estrategia de ramificación binaria en profundidad y se puede resumir como sigue:

En cada nodo se selecciona el ítem j con el máximo valor por unidad de peso de entre los ítems que aún no están seleccionados, y se generan dos nodos descendientes asignando a x_j respectivamente el valor 1 y 0. La búsqueda continúa desde el nodo asociado con la selección del ítem j (condición x_j=1). Y finalmente revisa las decisiones tomadas anteriormente mediante movimientos de retroceso. Un movimiento hacia adelante consiste en insertar el mayor número posible de nuevos ítems consecutivos en la solución actual, y un movimiento de retroceso consiste en borrar el más reciente ítem insertado de la solución actual. El Criterio de parada de este algoritmo, consiste en que el algoritmo finaliza cuando no se puede efectuar ningún retroceso.

2.2.2 El algoritmo dinámico con estados reducidos

Desde el punto de vista de la Programación Dinámica, la descomposición natural del PM consiste en los sub-problemas PM(m,c^l), con m ($1\leq m\leq n$) ítems, y con capacidad c^l, $0\leq c^l \leq$c. Sea

$$f_m(c^l) = max\left\{ \sum_{j=1}^{m} p_j x_j : \sum_{j=1}^{m} w_j x_j \leq c^l \right\} \qquad (6)$$
$$con \quad x_j \in \{0,1\},\ j = 1,\ldots, m.$$

el valor de la solución óptima de PM(m, c^l) .

Para resolver el PM la programación dinámica considera n etapas (igual al número de ítems) y calcula en cada una de ellas los valores dados por (6) usando la recursión clásica de Bellman-Dantzig siguiente:

$$f_m(c^l) = \begin{cases} f_{m-1}(c^l) \\ max\left(f_{m-1}(c^l),\ f_{m-1}(c^l - w_m) + p_m\right) \end{cases} \qquad (7)$$
$$para \quad c^l = w_m,\ldots, c$$

Las soluciones factibles asociadas a estos valores se denominan estados. La solución óptima del problema es el estado correspondiente a $f_n(c)$.

Horowitz y Sahni [2] propusieron la eliminación de aquéllos estados que no influyen en la determinación de los estados óptimos, llamados estados dominados. Dando origen así al algoritmo dinámico con estados reducidos (PD). Este algoritmo en cada etapa calcula y reserva los estados no dominados, los mismos que utiliza como valores de entrada para determinar los estados de la siguiente etapa. Como resulta evidente, el criterio de parada en este algoritmo no está netamente especificado, aunque es patente que el proceso finaliza cuando se ha obtenido el estado $f_n(c)$.

3 Criterios de parada difusos para el PM

Al igual que en muchos problemas NP-completos, en el problema de la mochila se admiten soluciones aceptables aunque no sean óptimas, para lo que pueden emplearse algoritmos heurísticos. Sin embargo aquí la alternativa que proponemos es la de flexibilizar los algoritmos exactos, introduciendo criterios de parada difusos, que finalmente proporciona una nueva clase de algoritmos heurísticos basados en reglas.

Para ilustrar el enfoque propuesto, consideramos un PM como (1), en el que para dar mayor sentido al razonamiento, suponemos que su dimensión es lo suficientemente considerable, como para que no sea posible el cálculo de la solución óptima del mismo. Es evidente que no toda solución puede ser aceptable, así como que el decisor siempre tendrá una idea imprecisa del rango de aceptabilidad sobre el valor final. Esta característica imprecisa, como resulta obvio, se representa perfectamente como un conjunto difuso, en lugar de cómo un intervalo.

Consecuentemente, asumimos que el conjunto difuso que representa la aceptabilidad propuesta por el decisor se define de la siguiente manera:

$$\mu(z) = \begin{cases} 0 & z < L_0, \\ f(t) & L_0 \le z \le U_0. \end{cases} \qquad (8)$$

donde $f(.)$ es una función continua no decreciente con valor entre [0,1], y L_0 y U_0 son cotas, inferior y superior respectivamente, del valor óptimo z^* de la solución óptima de PM, es decir, $L_0 \le z^* \le U_0$.

En general estas cotas tienen que estar dadas por el decisor. Pero si el problema es de gran dimensión, como aquí suponemos, lo más fácil es pensar que el decisor no sea capaz de precisarlas, por lo que puede ser una muy buena referencia como cota inferior L_0, el valor que pueda proporcionar cualquier algoritmo "greedy" aplicado al PM, y como cota superior U_0, la cota de Dantzig.

Sobre la base de las cotas y de la función de pertenencia, si el decisor considerara aceptable una solución factible cuyo valor z tenga un grado de pertenencia mayor que $\alpha \in (0,1)$ entonces el criterio de parada sería:

$$f(z) \ge \alpha \quad \text{ó} \quad z \ge f^{-1}(\alpha) \qquad (9)$$

Al introducir este criterio de parada en los algoritmos estudiados, estos se flexibilizarán de manera que se puede detener el proceso de cálculo cuando se obtenga la condición (9).

Una vez fijado el grado de aceptación α existen dos posibilidades, para el valor óptimo z^*:

a) Que sea mayor que $f^{-1}(\alpha)$, en cuyo caso se obtendrá una solución aceptable,
b) Que sea menor o igual que $f^{-1}(\alpha)$, en cuyo caso el criterio de parada difuso no tendrá sentido ya que el proceso nunca terminará, razón por la cual es necesario mantener los criterios exactos, y que el criterio difuso tenga carácter subsidiario.

Una forma de reducir la posibilidad de que se presente el caso (b) consiste en utilizar una función de pertenencia $f(.)$ cóncava, ya que entonces se cumple que

$$h^{-1}(\alpha) < f^{-1}(\alpha)$$

cuando h sea cóncava.

Una función cóncava que se puede utilizar en la definición de la función de pertenencia (8) puede ser de la siguiente forma:

$$f(t) = \sqrt[n]{\frac{t - L_0}{U_0 - L_0}} \qquad (10)$$

donde $n>1$, con lo que (9) se transforma en:

$$z \ge U_0 + (U_0 - L_0)\alpha^n \qquad (11)$$

Como ilustración, a continuación modificamos los criterios de parada de los algoritmos que introdujimos en la anterior sección.

3.1 Nuevo criterio de parada en el algoritmo HS

En el algoritmo de Horowitz-Sahni el criterio de parada consiste en detener el procedimiento cuando no existe ninguna posibilidad de retroceder. Al flexibilizar el algoritmo, además de mantener el criterio anterior incluimos la condición (9). Con este criterio de parada adicional, el algoritmo HS analizará cada vez que obtiene una nueva solución, si el valor correspondiente satisface o no a la condición (9), de ser positivo el proceso termina proporcionando la correspondiente solución aceptable.

3.2 Nuevo criterio de parada en el algoritmo PD

En este algoritmo el criterio de parada indica que el procedimiento se detiene cuando se obtiene el estado con el valor $f_n(c)$. Con este criterio, el procedimiento seguirá hasta el último estado de la última etapa, aunque la solución exacta se haya obtenido en las etapas anteriores o estados anteriores. Al introducir el criterio de parada difuso, en cada etapa y en cada estado, se analizará si se ha obtenido un valor aceptable o no; si es afirmativo, se detiene el proceso. Naturalmente, con este criterio no hay necesidad de resolver todas las etapas ni todos los estados.

4 Resultados experimentales

En esta sección presentamos los resultados de los experimentos computacionales realizados con problemas de la mochila de 1.000, 5.000, 10.000 y 50.000 ítems, cuyos valores y pesos comprendidos en el intervalo [1,1000], han sido generados aleatoriamente, y se ha utilizado la semi-suma de los pesos como capacidad de la mochila.

En cada uno de los algoritmos con criterios de parada difusos vistos en la sección anterior se considera un grado de pertenencia α=0.5, α=0.8 y α=1; este último caso, como es obvio, corresponde a la solución obtenida mediante el algoritmo clásico, sin el criterio de parada

difuso. En la tabla de los resultados se consignan el error cometido y el tiempo de ejecución para cada uno de los algoritmos de la sección anterior. El error se obtiene mediante la siguiente fórmula:

$$error = \frac{(Cota - Dantzig) - (sol.obtenida)}{cota\,de\,Dantzig} * 100$$

Por otro lado, para comparar los resultados que se han obtenido, dado el carácter aproximado de los algoritmos que hemos obtenido, recurrimos al algoritmo de Sahni [4], un algoritmo muy utilizado y bien conocido, que se corresponde con el siguiente esquema "greedy": El algoritmo trata de llenar la capacidad restante de la mochila, una vez introducido un conjunto M de k ítems. Para ello, en primer lugar, se generan todos los subconjuntos M posibles de k elementos para introducirlos en la mochila. Posteriormente rellena el espacio restante de la mochila. Por cuestiones obvias, el Algoritmo de Sahni suele usarse con k = 1, 2, 3.

En los criterios de parada difusos de los dos algoritmos revisados en la sección 2, cuyos resultados presentamos en esta sección, se ha utilizado la condición:

$$z \geq L_0 + (U_0 - L_0).\alpha^4$$

es decir, los criterios difusos están definidos mediante la función de pertenencia (8) con f(.) dada por (10) y n=4; L_0 es la solución respectiva obtenida mediante el algoritmo básico greedy y U_0 es la cota de Dantzig.

La tabla 1 muestra los resultados de los experimentos, tras haber utilizado los algoritmos con criterios de parada difusos de la sección 3, y paralelamente se consignan los resultados que proporciona el algoritmo aproximado de Sahni.

Como podemos observar en la tabla 1, el algoritmo de Horowitz y Sahni con criterio de parada difuso proporciona resultados con muy pocos errores y en tiempos también muy cortos, con respecto tanto a la solución exacta como a soluciones aproximadas dadas por el algoritmo de Sahni; esta última ventaja se hace más significativa cuando n (número de ítems) se incrementa. Por otro lado, se observa que el algoritmo dinámico con criterios de parada difusos proporciona una mejor aproximación (menor error) y en menor tiempo que el algoritmo de Sahni lo que junto con los resultados anteriores validan el enfoque propuesto.

5 Conclusiones

A la vista de todo lo anterior podemos concluir que la incorporación, a los algoritmos convencionales aquí considerados, de criterios de parada basados en reglas difusas, produce nuevos procedimientos de solución que, en las situaciones en que es difícil la obtención de soluciones óptimas, permiten resolver los problemas en cuestión a satisfacción del decisor. La generalidad y robustez del método que se deriva del uso de los sistemas basados en reglas, mas que contrastado en muchos contextos, hacen prever la adecuación del enfoque aquí descrito a una gran variedad de situaciones.

Referencias

[1] F. Herrera y J.L. Verdegay. Fuzzy control rules in optimization problems, *Scientia Iranica*, Vol 3, Nos.1,2,3. (1996) 89-96.

[2] E. Horowitz y S. Sahni. Computing partitions with applications to the knapsack problem, *Journal of ACM* 21 (1974) 277-292.

[3] S. Martello y P. Toth. *Knapsack Problems*, John Wiley and Sons 1990.

[4] S. Sahni. Approximate algorithms for the 0-1 knapsack problem, *Journal of ACM* 22 (1975) 115-124.

[5] J.L. Verdegay. Fuzzy Optimization: models, methods and perspectives; *6th IFSA-95 World Congress*. Sau Paulo - Brazil, 1995, pp. 39-7.

[6] E.R. Vergara. *Nuevos Criterios de Parada en Algoritmos de Optimización*, Dpto. de CC. de la Computación e I.A. (Tesis Doctoral), 1999.

Tabla 1. Resultados de los experimentos, mediante HS y PD con criterios de parada difusos.

n	α=1 (exacto)				α=0.5				α=0.8				Algoritmo de Sahni			
	HS		PD		HS		PD		HS		PD		k=2		k=3	
	Error	tiemp	error	tiemp	error	tiemp	Error	tiemp	Error	tiemp	Error	Tiemp	error	Tiemp	Error	Tiemp
1.000	.00192	0.044	.00192	0.24	.00517	0.000	.00436	0.19	.00395	0.000	.00382	0.23	.00325	0.11	.00247	0.23
5.000	.00009	1.462	.00009	3.77	.00046	0.000	.00027	3.07	.00031	0.022	.00024	3.10	.00023	2.54	.00017	4.58
10.000	.00002	3.720	.00002	12.19	.00011	0.010	.00006	10.18	.00006	0.012	.00003	10.25	.00029	8.80	.00027	13.97
50.000	.00000	21.838	.00000	223.94	.00002	0.096	$4/10^7$	218.35	.00001	0.136	$3/10^7$	218.37	$3/10^7$	212.48	.00000	257.76

LINGUISTIC VARIABLES DETERMINATION USING FUZZY CLUSTERING

Antonio Flores-Sintas[1]
Dpto. de Nuevas Tecnologías
Caja Ahorros de Murcia
Gran Vía,23
30005-Murcia
e-mail: aflores@cajamurcia.es

José M. Cadenas **Fernando Martín**
Dpto. Informática, Inteligencia Artificial y Electrónica
Facultad de Informática
Universidad de Murcia
30071-Espinardo.Murcia
email: jcadenas@dif.um.es email: fmartin@dif.um.es

ABSTRACT

The fuzzy sets defining the linguistic variable values can be seen as a fuzzy partition of the linguistic variable. The membership functions obtained using fuzzy clustering algorithms are defined with respect to the group prototypes, and they cannot be used to define the linguistic variable values. We introduce several criteria to pass from the clustering membership functions to the linguistic variable value membership functions. The data have been taken from the financial sector.

Keywords: membership functions, cluster analysis, fuzzy c-means.

1. INTRODUCTION.

"The clients must guide the actions of the financial companies and these must adjust the offer of products and services to satisfy the needs of the clients. The marketing policy must be founded on strategic lines of action just as to classify the clients, to make blocks of products, to personalise the financial products and to offer products and services of quality. However, it is not easy to classify the clients, and it is more difficult to do it if we want classify clients belonging to the home economies sector because this is a wide and heterogeneous set. The classification must be pointed toward the market. For this reason, among the possible criteria to be used, it is necessary give a greater relevance to the advantages searched by the clients when they are related with the financial companies, and to the products and services used by the clients".(translated from [2])

The financial sector is highly computerised and it is a direct sales sector. As a consequence, the information about the consumption of the clients is very high. However, the number of products and services offered is, as any direct sales sector, much larger than the average number of products and services consumed by client. This fact makes difficult the application of conventional techniques of classification in the features global space (products and services) [1]. On the other hand, the bank clerks daily realise a classification of their clients using linguistic variables of the type: he has a medium debit balance, a low loan, a high investment fund, etc. It seems reasonable to use the same criterion and to realise one-dimensional classifications to establish the profile of the clients. In [6] we have realised an approximation to this problem, and this contribution is a sequel of [6].

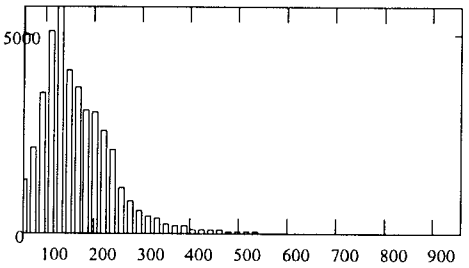


Figure 1. Number of clients/feature value

To determine the profile of a client we can use any countable feature providing information about the client. In this sense as example, the interest payment to a client can be used as a feature although it is not a product or a service offered but a financial cost. We have detected two features types distributed as figures 1 and 2. The salaries payment or the charges of the home insurances are as figure 1. The interest payment or the charges of the life insurances are as figure 2. We can see, for both cases, that there are not natural groups and that the distributions are very not homogeneous (the data are taken from Caja de Ahorros de Murcia). Referring to the first problem, we are not interested in detecting natural groups but in dividing a one-dimensional sample in a number of groups

[1] **Corresp**onding author

adjusting them as to the values of a linguistic variable (very-low, low, medium, high, very-high). In the contribution we will use five groups corresponding to the five previous linguistic values. Referring to the second problem we cannot use the Euclidean distance because the distributions are very not homogeneous ([3],[4],[5]). We will use the Mahalanobis distance and the Fuzzy C-Means algorithm (FCM), modified in [4] which is based on [3], to find a good partition. However, the grades of membership found by the FCM algorithm are referred to the detected prototypes and they cannot be used as grades of membership to the values of the linguistic variable. In this paper, we introduce several criteria to pass from the membership functions found by the FCM algorithm to the membership functions defining the values of the linguistic variable.

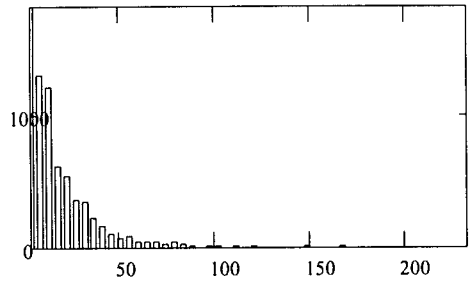


Figure 2. Number of clients/feature value

The FCM algorithm obtains the prototypes by Picard iterations from any hard partition, minimising a within-group sum of squared errors objective function. The FCM algorithm modified in [4] (it can also be in [6]) differences between the grades of membership and the probabilities of membership. In this paper we only are interested in the grades of membership, which are given by the following expression:

$$\mu_{xk} = \sqrt{\left|C_k^{-1}\right|} \Big/ \left(1 + d_{xk}^2\right)^{3/2}$$

μ_{xk} being the grade of membership of the x-pattern to the k-group; C_k being the k-group fuzzy covariance matrix, and $d_{xk}^2 = (x - v_k)^T C_k^{-1} (x - v_k)$ being the Mahalanobis distance between the x-pattern and the v_k-prototype. Notice that the range of the membership function is different from the unit interval. This is not contradictory with the fuzzy set theory [7].

Figures 3 and 4 represent the membership functions obtained by using the FCM algorithm corresponding to the samples whose distributions are shown in figures 1 and 2.

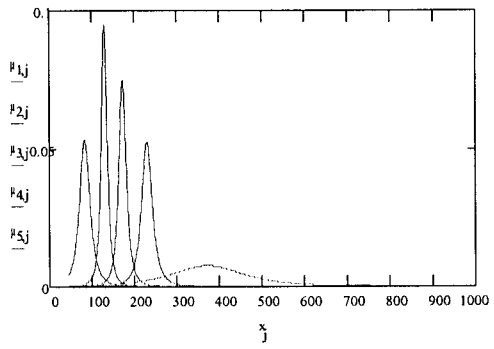


Figure 3. Membership functions (FCM) corresponding to figure 1.

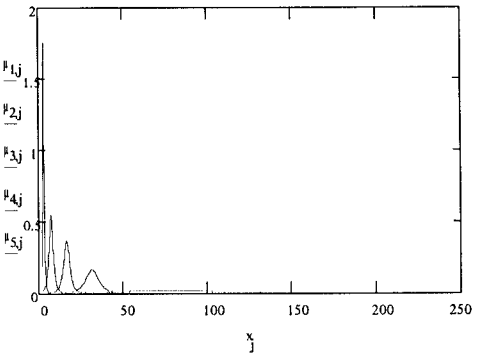


Figure 4. Membership functions (FCM) corresponding to figure 2.

2. MEMBERSHIP FUNCTIONS TO THE LINGUISTIC VARIABLE VALUES.

To distinguish between the membership functions calculated by the FCM algorithm and the membership functions of the linguistic variable values, we are going to represent by the symbol η the second membership functions, maintaining the symbol μ for the first ones. η_{xk} will represent the grade of membership of the x-pattern of the linguistic variable value of the k-group,

being the 1-group=very-low, 2-group=low, etc., or well, the selected label of each group.

As we are working in one dimension, we can define the following criteria to determine the grades of membership of the patterns of the linguistic variable values:

1. All the prototypes obtained from the FCM algorithm belong, with all security, to the linguistic variable value correspondent to the group represented by each prototype (for example, 1-group to very-low, 2-group to low, etc).

2. The patterns placed at the left of the v_1-prototype belong, with all security, to the first linguistic variable value (for example, very-low), and the patterns placed at the right of the v_c-prototype belong, with all security, to the last linguistic variable value (for example, very-high).

3. One pattern can have, at most, two grades of membership to the linguistic variable values different from zero. For example, if the x-pattern is placed between the v_1 and the v_2 prototypes, the grades of membership of the x-pattern to the very-low and the low linguistic variable values will be larger than zero, but the grades of membership to the others linguistic variable values will be zero.

Mathematically, the first and the second criteria can directly be expressed. Let us see how we can express the third criterion: Figure 5 represents a hypothetical membership function obtained with the FCM algorithm respect of a hypothetical v_j-prototype ($v_j = 10$ in the figure). Let us assume $v_{j+1} = 20$ in the figure. Let us take an element of the sample (x) placed on the axe of abscises and the line separating the grey painted area from the dark painted area.

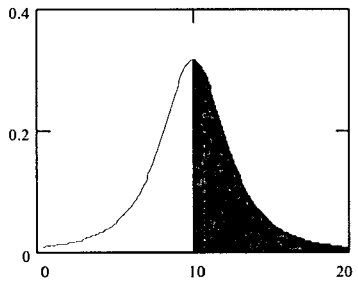


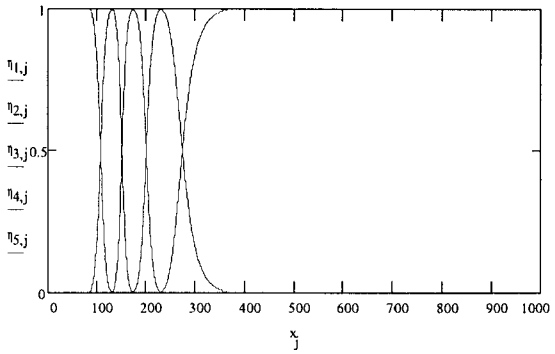
Figure 5

Let us define the following magnitude:

$$w_{xj} = \frac{dark\ area}{grey\ area + dark\ area}$$

When $x \to v_{j+1}$, $w_{xj} \to 0$. When $x \to v_j$, $w_{xj} \to 1$. Moreover, using a geometrically symmetrical criterion for the patterns placed at the left of the prototypes, and defining $w_{xj} = 0 \quad if \quad \left(x < v_{j-1} \ or \ x > v_{j+1}\right)$, the magnitude w_{xj} satisfies the second and the third criteria to determine the membership functions of the values of the linguistic variables, but taking only into account one group. We need calculate the previous magnitude as relative to determine the membership functions searched for.

Figure 6. Membership functions of the linguistic variable values corresponding to figure 1.



The μ-membership functions using one-dimensional samples are expressed as:

$$\mu_{xk} = \frac{\sqrt{\sigma_k^{-1}}}{\left(1 + \sigma_k^{-1}\left(x - v_k\right)^2\right)^{3/2}}$$

σ_k being the k-group fuzzy variance. The areas under the μ-membership functions are very easy to calculate because:

$$\int \frac{\sqrt{b}}{\left(1 + b(z-a)^2\right)^{3/2}} = \sqrt{b}\,\frac{z-a}{\sqrt{1 + b(z-a)^2}} + cte$$

Let us define the following function:

$$\mathrm{A}(z;a,b) = \frac{z-a}{\sqrt{1 + b(z-a)^2}}$$

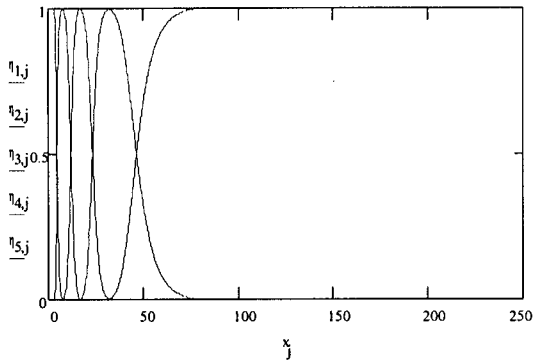


Figure 7. Membership functions of the linguistic variable values corresponding to figure 2.

The w-magnitudes are calculated as follows:

$$if\ x < v_1 \Rightarrow w_{x1} = 1, w_{xj} = 0, 1 < j \le c$$

$$if\ x > v_c \Rightarrow w_{xc} = 1, w_{xj} = 0, 1 \le j < c$$

$$if\ v_1 \le x \le v_c \Rightarrow (for\quad 1 \le j \le c)$$

$$\Rightarrow \left\{ \begin{array}{l} if\ v_{j-1} \le x \le v_j \Rightarrow w_{xj} = \dfrac{A(x; v_j, \sigma_j^{-1}) - A(v_{j-1}; v_j, \sigma_j^{-1})}{A(v_j; v_j, \sigma_j^{-1}) - A(v_{j-1}; v_j, \sigma_j^{-1})} \\ if\ v_j \le x \le v_{j+1} \Rightarrow w_{xj} = \dfrac{A(v_{j+1}; v_j, \sigma_j^{-1}) - A(x; v_j, \sigma_j^{-1})}{A(v_{j+1}; v_j, \sigma_j^{-1}) - A(v_j; v_j, \sigma_j^{-1})} \\ if\ x < v_{j-1}\ and\ x > v_{j+1} \Rightarrow w_{xj} = 0 \end{array} \right\}$$

The w_{xj} are defined in such a way that they could be considered as absolutes membership probabilities to the linguistic variable values. To calculate the relative probabilities, we are going to use the Bayes theorem. The probability densities will be $w_{xk} / \sum_{z \in X} w_{zk}$. We need know the a priori probabilities to apply the Bayes theorem. In [4] and [5], we have given to the membership functions calculated using the FCM algorithm the meaning of densities. Figures 3 and 4 are with this interpretation. The largest density of each group is at each prototype, and it decreases when we go away from the prototype. This is because the prototype represents the group. Increasing the number of groups, the densities at the prototypes will be on a curve representing the sample density. The "mass" belonging to each group will be $\sum_{x \in X} \mu_{xk}$. We take the a priori probabilities as proportional to $\sum_{x \in X} \mu_{xk}$. In consequence, the a posteriori probabilities will be:

$$\eta_{xk} = \frac{w_{xk}}{\sum_{z \in X} w_{zk}} \sum_{y \in X} \mu_{yk} \Big/ \sum_{j=1}^{c} \frac{w_{xj}}{\sum_{z \in X} w_{zj}} \sum_{y \in X} \mu_{yj}$$

The a posteriori probabilities are the membership functions searched for. Notice that a probability is a grade of membership, but vice versa is not correct. Figures 6 and 7 represent the membership functions of the linguistic variable values corresponding to figures 3 and 4.

Using similar criterions, the membership functions of the linguistic variable values can be determined when we use another classification technique.

3. CONCLUSIONS.

We have obtained the fuzzy sets defining the linguistic variable values from one-dimensional samples using fuzzy clustering techniques. As a consequence, the fuzzy sets are adjusted to the sample distributions. The results of this contribution can be used to apply the fuzzy inference techniques as in the financial sector as another sector. This implies the previous application of the knowledge acquisition techniques. We have applied the results of this contribution to the sale possibility assessment of financial products, which will be the object of another contribution.

4. REFERENCES.

[1] Duda P., Hart P., Pattern Classifications and Scene Analisys, Wiley, New York, 1973.

[2] Egea Krauel C, Análisis Estratégico del Sector de Cajas de Ahorro en España, Tesis Doctoral, Universidad Autónoma de Madrid, 1988.

[3] Flores-Sintas A.,Cadenas J.M.,Martin F., A Local Geometrical Properties Application to Fuzzy Clustering, Fuzzy Sets and Systems, 100 (1998) 237-248.

[4] Flores-Sintas A.,Cadenas J.M.,Martin F., Membership Functions in the Fuzzy C-Means Algorithm, Fuzzy Sets and Systems, 101 (1999) 49-58.

[5] Flores-Sintas A.,Cadenas J.M.,Martin F., Partition Validity and Defuzzification, Fuzzy Sets and Systems, to be published.

[6] Flores-Sintas A., Cánovas J., García J., Aproximación a la determinación de perfiles de clientes en entidades financieras utilizando técnicas fuzzy, Actas del VIII Congreso Español sobre Tecnologías y Lógica Fuzzy, (Pamplona, Spain, 1998) 431-435.

[7] Zadeh L., Fuzzy Sets, Inform. and Control, 8-3 (1965) 338-353.

Fuzzy Reasoning In K-Means Classification Method

Irene DIAZ, Manuel VELASCO and Jose M. MOLINA

Departamento de Informática
Universidad Carlos III of Madrid, Butarque, 15
28911 Leganés, Madrid, España

Summary

Domain analysis tries to reuse software in an effective way. New methodologies are starting to be able to automate the process, in different degrees, with the construction of a domain model for each problem. The general process is divided into several phases. One of the most difficult tasks is the generation of the relationships which have to be defined between the components in the domain. In this paper the use of fuzzy logic and a statistical classification method in order to get the semantic relationships for each pair of components is presented.
Keywords: Domain analysis, Classificaton, Fuzzy System.

1 INTRODUCTION

Information classifying techniques usually use clustering algorithms to order the information. Each of those methods must measure distances between two vectors (that represent the information [8]). These distances are used to decide if a given vector belongs to a cluster or not. As the nature of the problem is often inexact and subjective, the use of a traditional distance does not provide successful results because such distances do not take account of all semantic information known about the data. Then, the use of a knowledge-based system to solve the problem allows us to take advantage of the semantic information that a domain provides us with. So, this distance should improve results obtained from traditional distances. These kind of fuzzy distances have been used, for instance, in the processing of stereoscope images to calculate distances among points of different images [3].
In this paper we use fuzzy reasoning in the subject of construction and validation of a methodology for automatic construction of domains [5]. Domain analysis was introduced in [6] as "An activity of identifying objects and operations of a class of similar systems in a domain of the specific problem". A process of global domain construction is shown in [8]. That method has five phases which are *Plan of the project and domain characterisation, domain components acquisition, data analysis, classification of components, and evaluation of the domain model.*
Classification phase tries to cluster the knowledge considered as similar. Once the knowledge (represented as terms or 'descriptors') has been clustered at a first level, common characteristics more relevant of each cluster are abstracted again and each cluster is divided into other clusters. The result of that process organised the knowledge in hierarchies.
In order to classify the information, statistical, neural and bibliometric methods are used to get these relationships. In this work, a statistical method is presented, the K-means. To sum up, when the process is beginning, we have a set of vectors $\{x_1, ..., x_m\}$ which represent information about each term. Each component x_{ij} of a vector x_i means the number of occurrences of term i in document j. Our purpose is to obtain a set of classes $\{a_1, ..., a_n\}$ in which these vectors are grouped. The problem is how these vectors are divided along the classes and how many classes will be.
Most of these methods are based on minimisation on some index. This process implies the use of a distance. In this work, it is proposed the use of a fuzzy distance [4], result of applying a fuzzy rule based system, which improves results obtained with an usual distance.

2 CLASSIFICATION PROBLEM

The problem is the classification of information over a particular domain [1],[5].To classify the information clustering analysis is used in this work. Clustering can be defined as an unsupervised process to classify objects. The distribution of the objects into the different

classes is not well known; but a vectorial representation of the objects is available $\{x_1, ..., x_p\}$. From this set of vectors, the classification process try to achieve the set of classes that group them $\{a_1, , a_n\}$. The problem is that neither the distribution nor the number of classes is known before the process. However the structure of the features vector is well defined (built up with the number of occurrences of the current candidate in each document).
Therefore, the problem consists in getting groups from the set of vectors in function of the similarities found between them.
Clustering is carried out according to the pattern similarity, by means of the used distance. The introduction of a fuzzy distance is based on the need of gathering as semantic information as possible.
The classification algorithm used in this work is K-Means. This algorithm is one of the most popular for clustering analysis. There are several algorithm versions, and one of the most efficient is the Anderberg's convergence algorithm [2]. It is an iterative and more formalised algorithm, in front of other excessive heuristic, for example the max-min algorithm. It is a quick and efficient algorithm which tries to minimise an index, based in adding all the quadratic Euclidean distances (fuzzy distances in this case) of all the members in a cluster to the cluster centre. It needs to know the k number that defines the number of cluster in the collection. If the number of classes is unknown then it is a trouble. One efficient solution consists in defining the number of classes using some user-defined parameters. If a collection is available with N training patterns, represented by x_i vectors, the centre (centroid) in this set of vectors is defined by

$$\frac{1}{N}\sum x_i \tag{1}$$

The way the algorithm works consists in moving each vector to the cluster which centre is the closest to the selected vector, and later the algorithm calculates again the centre in each cluster. Algorithm convergence depends on the number of classes. This algorithm works quite well when the number of classes is known or estimated.

3 FUZZY TERM MEASURE

Let us consider multi-input, single-output fuzzy systems mapping $X \longrightarrow Y$, where $X \subset \Re^n$ and $Y \subset \Re$. A basic configuration of fuzzy system is composed of a fuzzifier, a rule system and a defuzzifier [6]. The fuzzifier performs a mapping from the observed crisp input space to the fuzzy systems defined.
A rule base of a fuzzy system consists on a number of fuzzy rules , expressed as IF-THEN form. For the fuzzy system we consider suppose you have a set of N rules

$$R^k : IF\ x_i^k\ is\ X_i^k\ AND\ ...\ AND\ x_n^k\ is\ X_n^k$$

$$THEN\ y\ is\ Y^k \tag{2}$$

where X_i^k and B^k, $i = 1, ..., n$ and $k = 1, ..., N$ are linguistic terms characterized by fuzzy membership functions. Now we define a fuzzy distance which allow us to classify the set of vectors $\{x_1, ..., x_m\}$ into n classes a_i, $i = 1, ..., n$. Each vector means the number of occurrences of a term in each document, i.e., there are vectors $x_i = \{x_{ij}\} j = 1, ..., n$, where x_{ij} means number of times that the term i appears on document j. Using an expert system to build the fuzzy distance adequate to the problem, it is possible to use some information that should be overlook if an usual distance is used. This information is essential to solve the problem successfully. The objective is to know if two terms are related in some sense, that is, if two terms are synonyms, one is generic to other one, or if they treat about the same theme. The proposed multi-input single-output fuzzy system has four input variables and one output variable [1].These are briefly explained in the next subsections.

3.1 INPUT VARIABLES

3.1.1 NOT ZERO COINCIDENCES

It seems that terms that belong to same documents treat about more or less the same subject. Then, it is significant that a vector's component would be zero or not because this fact indicates that a term is, or is not, in a document. We define "Not zero coincidence"(nzc) input variable as follows. Let x_i and x_j be two vectors, It is said that one of their component k, coincide iff

$$x_{ik} \neq 0 \neq x_{jk} \tag{3}$$

The number of these coincidences will be an input to the system. This input treats to group those terms that are in the same documents. That fact is reasonable because two terms in different documents will have less likelihood to be related in some sense. Note that it is not very important the number of documents in which a term should be. The really important things are two, the first one is that two terms are, or not, in the same document and the second one how many documents are simultaneously two terms in.

3.1.2 NOT COINCIDENCE

It is also important to know if two terms are not simultaneously in a document, because if two terms are in different documents, the relationship between

them will probably be small. We define "Not coincidence"(nc) input variable as follows. Let x_i and x_j be two vectors, It is said that one of their component k, coincide iff

$$x_{ik} \neq 0 \, AND \, x_{jk} = 0 \, OR \, x_{ik} = 0 \, AND \, x_{jk} \neq 0. \quad (4)$$

Summing over k we obtain the input value for this input variable,that is, we obtain the number of documents in which the two terms under study are not simultaneously.

3.1.3 MAXIMUM DIFERENCE IN THE NUMBER OF OCCURRENCES

This input variable (no) tries to weight the difference between the number of times that two vectors appear in the corpus. So for each corpus the maximum number of occurrences for a term is estimated as follows. *Given two vectors(representing terms) x_i and x_j, the maximum difference in the number of occurrences for those terms are*

$$max_{k=1,\ldots,n}|x_{i_k} - x_{j_k}| \quad (5)$$

3.1.4 NUMBER OF DOCUMENTS

The number of documents (nd) is a variable used to weight the relevance of the size of the corpus used. In fact, the number of documents is very important for the classification, because the information does not grow linearly with the growing of the number of documents.

3.2 OUTPUT VARIABLE: DISTANCE

The output is a value that indicates whether two vectors are near or not. The relationship between two descriptors is inversely proportional to the obtained value. To calculate output fuzzy sets, the bound of the value range must be known.

3.3 SYSTEM RULES

We have used three linguistic labels

Small, Medium, High

for each input variable and five for the output variable

Very Small, Small, Medium, High, Very$h$igh .

Now we present some of the rules we use:

IF nzc is S AND nc is S AND no is S AND nd is A

THEN distance is S

IF nzc is S AND nc is S AND no is H AND nd is S

THEN distance is Vh

IF nzc is S AND nc is M AND no is M AND nd is S

THEN distance is H

IFnzc is M AND nc is SAND no is S AND nd is S

THEN distance is S

IFnzc is M AND nc is S AND no is S AND nd is M

THEN distance is S

IFnzc is M AND nc is S AND no is SAND nd is H

THEN distance is Vs

IFnzc is M AND nc is S AND no is M AND nd is S

THEN distance is M

IFnzc is H AND nc is S AND no is H AND nd is H

THEN distance is Vs

IFnzc is H AND nc is M AND no is S AND nd is S

THEN distance is Vs

IFnzc is H AND nc is M AND no is S AND nd is M

THEN distance is S

IFnzc is H AND nc is M AND no is S AND nd is H

THEN distance is S

4 RESULTS

In the experiments fulfilled to contrast the results using the euclidean distance in k-means or the fuzzy distance in K-means algorithm presented here, we have made several experiments varying the number of documents (that is, the dimension of the vectors)in the corpus. Here we present one of those experiments.
We have selected only two documents in this example so that the number of terms is low (212).
The number of terms that is represented by the same vector is shown in table 1. In table 2 we can see in the

Table 1: Number of terms represented by the same vector

VECTOR	NUMBER OF TERMS
(1, 0)	41
(2, 0)	16
(3, 0)	5
(4, 0)	1
(5, 0)	3
(6, 0)	2
(8, 0)	1
(10, 0)	2
(15, 0)	2
(0, 1)	24
(0, 2)	3
(0, 6)	1
(0, 7)	1
(1, 1)	2
(2, 2)	1
(3, 3)	1
(2, 1)	1
(3, 2)	1
(4, 1)	1

Table 2: K-Means Classification versus K-Means with Fuzzy Reasoning Classification

VECTOR	K-Means Class	FK-Means Class
(1, 0)	1	1
(2, 0)	3	1
(3, 0)	3	1
(4, 0)	3	3
(5, 0)	3	3
(6, 0)	3	3
(8, 0)	5	3
(10, 0)	5	3
(15, 0)	5	5
(0, 1)	1	2
(0, 2)	1	2
(0, 6)	2	2
(0, 7)	2	2
(1, 1)	1	4
(2, 2)	4	4
(3, 3)	4	4
(2, 1)	4	4
(3, 2)	4	4
(4, 1)	3	4

second column the final result obtained by k-means and in the third one, the final result using k-means with fuzzy reasoning.

As result of this research we can conclude the classification that provide us K-Means with Fuzzy Reasoning is better for the purposes of this research because that classification is according with the semantic of the terms.

References

[1] Díaz,Velasco and Molina,*An application of fuzzy logic to domain analysis*,Proc. EUROFUSE-SIC'99/Pag 434-439. Budapest, May 1999.

[2] Gallant, *Neural Network Learning and Expert Systems* The MIT Press. 1995.

[3] Jiménez,Molina and Casar,*A Fuzzy Reasoning System for Binocular Scene Matching*, Proc. IEEE Workshop on NonLinear Signal/Image Processing, Pag 309-312. Greece, June 1995.

[4] Klir,Folger,*Fuzzy Sets, Uncertainty and Information*, Prentice-Hall, 1992.

[5] Lloréns,Velasco,Moreiro,Martínez and Amescua, *Automatic Domain Analysis using Thesaurus Structures*,Accepted in Journal of the American Society of Information Science, Marzo 1998.

[6] Neighbors, *Software Construction using Components*,Ph. D. Thesis. Department of Information and Computer Science. University of California. 1981.

[7] Rutkowska and Nowicki, *Constructive and Destructive Approach to Neuro-Fuzzy Systems*.Proc. EUROFUSE-SIC'99/Pag 100-105. Budapest, May 1999.

[8] Velasco,Martínez,Lloréns and Amescua, *Automatic Domain Analysis: Generation of Domain Representations*. IT-KNOWS 98. Viena. September 1998.

A SURVEY OF METHODS FOR EVALUATING QUANTIFIED SENTENCES

Miguel Delgado, Daniel Sánchez & Maria Amparo Vila
Department of Computer Science and A.I.
University of Granada, Granada, Spain
e-mail: {delgado, daniel, vila}@decsai.ugr.es

Abstract

The evaluation of quantified sentences is used for solving several problems. Most of the methods proposed are not satisfactory because they do not verify some intuitive properties. In this paper we propose an extension of both possibilistic and probabilistic methods, based on the Sugeno and the Choquet fuzzy integrals respectively, for the evaluation of type II sentences, the most general kind of sentences. These methods have good properties, and they are shown to be better than existing ones.

Keywords: Quantified sentences, quantifiers, fuzzy integrals.

1. Introduction

Quantified sentences are assertions about the number or percentage of objects that verify a certain property. They have applications in many fields such as flexible database querying and expert systems among others (see [8] for a brief review). These sentences are classified into two classes, called type I sentences and type II sentences. A type I sentence is a sentence of the form "Q of X are A", where $X=\{x_1,...,x_n\}$ is a finite set, Q is a linguistic quantifier and A is a fuzzy property defined over X. One example of such a sentence is "Most of the students are young". In this case, X is a finite set of students, the property A is "young" and Q is the quantifier "Most". A type II sentence can be described in general as "Q of D are A", where D is also a fuzzy property over X. One example is the sentence "Most of the young people are tall". Obviously, type I sentences are a special case of type II sentences where D=X.

The evaluation of quantified sentences tries to obtain an accomplishment degree in the real interval [0,1] for the sentence. In [5] we propose a set of intuitive properties that the evaluation must verify. According to those properties, we show that some existing evaluation methods are not appropriate in some cases. The case of the evaluation with non-coherent quantifiers is very significant for both type I and type II sentences. A fuzzy quantifier defines a fuzzy quantity (absolute quantifier) or a fuzzy percentage (relative quantifier). Absolute quantifiers are defined as fuzzy sets over the nonnegative integers, while relative quantifiers are fuzzy sets over the real interval [0,1]. A quantifier Q is said to be coherent if it is non-decreasing and Q(0)=0 and Q(1)=1. The only method that works with non-coherent quantifiers is Zadeh's method, but this method has the drawback that it is very strict when evaluating crisp quantifiers such as "exists" and "forall". These problems are sufficient to justify the search for new methods with better properties. Nevertheless there are other properties that are not verified by some of the existing methods.

In our opinion, the best available methods before our work are based on Sugeno and Choquet fuzzy integrals, although they work only with coherent quantifiers. The latter is equivalent to the method of Yager based on the OWA operator ([9]). They can be interpreted as a possibilistic and a probabilistic approach respectively. In [2] the relation among them is studied, but none of them is preferred. One of our objectives is generalizing these methods for working with coherent quantifiers. Another objective is generalizing these methods for evaluating type II sentences, and studying the relation among them in this case to find some criterion, if possible, to choose one of the methods.

2. Type I sentences

2.1. Previous methods

An early method for evaluating type I sentences was introduced by Zadeh [11] to be

$$Z_Q(A) = Q\left(\frac{P(A)}{|X|}\right)$$

for relative quantifiers, where P(A) is the power of A defined as

$$P(A) = \sum_{i=1}^{n} A(x_i)$$

This method is very strict when evaluating crisp

quantifiers. One example is the case of the quantifiers "exists" and "all", that can be defined as

$$\forall(x)=\begin{cases}1 & x=1\\ 0 & x<1\end{cases} \qquad \exists(x)=\begin{cases}0 & x=0\\ 1 & x>0\end{cases}$$

Then, the evaluation of the sentences is

$$Z_{\forall}(A)=\begin{cases}1 & A=X\\ 0 & A\neq X\end{cases} \qquad Z_{\exists}(A)=\begin{cases}0 & A=\varnothing\\ 1 & A\neq\varnothing\end{cases}$$

Other existing methods for evaluating type I sentences are based on the Sugeno and the Choquet integral (see [1]). They are not strict methods, but both of them are restricted by definition to the case of monotonic quantifiers. The method based on the Sugeno integral is defined as

$$S_Q(A)=\max_{1\le i\le n} min(Q(i/n),b_i)$$

where b_i is the i-th greater value of $A(x_i)$, $i\in\{1,...,n\}$. The method based on the Choquet integral is defined as

$$C_Q(A)=\sum_{i=1}^{n} b_i*(Q(i/n)-Q((i-1)/n)$$

Example 1 shows that S_Q is not an appropriate method when dealing with non-coherent quantifiers. Example 2 shows that C_Q has the same problem.

Example 1: Let $X=\{x_1, x_2, x_3\}$ and $A=\{1/x_1, 1/x_2\}$ and let $Q(0)=0.5$, $Q(1/3)=1$, $Q(2/3)=0.5$, $Q(1)=0$. The quantifier Q can be linguistically interpreted as "approximately 1/3". Then we have the evaluation $S_Q(A)=\max\{\min(1,1), \min(0.5,1), \min(0,0)\}=1$. But A is crisp, $|A|=2$ and $|X|=3$, so the expected result of the evaluation was $Q(2/3)=0.5$.

Example 2: Let X and Q be as in example 1, and let $A=\{1/x_1\}$. Now we have $|A|=1$, so the expected result is $Q(1/3)=1$. Instead, C_Q gives a value $C_Q(A)=0.5$.

2.2. Our methods

In [3] we introduced an extension of the method C_Q, called GD, to be

$$GD_Q(A)=\sum_{i=0}^{n}(b_i-b_{i+1})*Q(i/n)$$

In the same paper we show that if Q is coherent, then $GD_Q(A)=C_Q(A)$. This extension can be used with non-monotonic quantifiers and it has good properties.

The results obtained for the examples 1 and 2 are the expected results. In general, it is shown in [5] that if A is crisp then $GD_Q(A)=Q(|A|/|X|)$. Moreover, the evaluation method is not strict in the sense of Zadeh's method, i.e. we can obtain a value in (0,1) even if the quantifier is crisp.

In [4] we introduced an extension called ZS of the method based on the Sugeno integral, to be

$$ZS_Q(A)=\max_{\alpha\in M(A)} min\left(\alpha, Q\left(\frac{|A_\alpha|}{|X|}\right)\right)$$

where

$$M(A)=\{\alpha\in\,]0,1] \mid \exists x_i\in X \text{ such that } A(x_i)=\alpha\}\cup\{1\}$$

We show in [4] that if Q is coherent then $ZS_Q(A)=S_Q(A)$. We also show that this method gives the expected result in the case of A being crisp, and it is not strict.

3. Type II sentences

3.1. Previous methods

Zadeh's method was introduced in [11] to be

$$Z_Q(A/D)=Q\left(\frac{P(A\cap D)}{P(D)}\right)$$

This is also an strict method with crisp quantifiers, for example for the quantifiers "exists" and "forall" we have

$$Z_{\forall}(A/D)=\begin{cases}1 & D\subseteq A\\ 0 & otherwise\end{cases}$$

and

$$Z_{\exists}(A/D)=\begin{cases}0 & A\cap D=\varnothing\\ 1 & otherwise\end{cases}$$

Yager introduced another method in [10] to be

$$Y_Q(A/D)=\sum_{i=1}^{n} w_i c_i$$

where c_i is the i-th largest value of membership to the fuzzy set $D'\cup A$, where D' is the standard complement of D, and

$$w_i=Q(S_i)-Q(S_{i-1}) \quad i\in\{1,...,n\}$$

and

$$S_i = \frac{1}{d}\sum_{j=1}^{i} e_i \quad and \quad d = \sum_{k=1}^{n} e_k$$

and e_k is the i-th smallest value of membership to D and S_0=0.

This method has several drawbacks. First, it is defined only for coherent quantifiers. The following examples show two more problems of the method.

Example 3. Let us consider A={1/x_1, 0/x_2} and let D={0/x_1,0.1/x_2}. Also let Q=∃. We have A∩D=∅, so clearly the percentage of objects of D that are in A is 0, and hence the evaluation of the sentence "Q of D are A" should give ∃(0)=0 (see [5]), but the method gives $Y_{\exists}$(A/D)=0.9 instead.

Example 4. Let us consider A={1/x_1, 0.9/x_2} and D={1/x_1, 0.5/x_2} and Q(x)=x. Clearly D⊆A and D is normalized, so the percentage of objects of D that are in A is 1, and hence the evaluation should be Q(1)=1 (see also [5]). Instead of the expected result, the method gives the result Y_Q(A/D)= 0.93. The same problem arises if Q=∀. The expected result is ∀(1)=1, but the method gives the result $Y_{\forall}$(A/D)=0.9.

3.2. Our methods

The previous evaluation methods for type II sentences are less satisfactory than the previous methods for type I sentences. We have generalized our methods GD and ZS to the case of type II sentences to obtain methods with good properties.

First we introduced in [4] the method ZS for type II sentences to be

$$ZS_Q(A/D) = \max_{\alpha \in M(A/D)} \min\left(\alpha, Q\left(\frac{|(A\cap D)_\alpha|}{|D_\alpha|}\right)\right)$$

where M(A/D)=M(A∩D)∪M(D). We assume D is a normal fuzzy set. If it is not, we normalize D in the usual way and we apply the same normalization factor to the fuzzy set A∩D (so, at least conceptually, we are not changing the relative cardinality of D with respect to A).

This method has several good properties as we show in [5]. First, if A and D are crisp, then ZS_Q(A/D) is the expected result Q(|A∩D| / |D|). Second if D⊆A then ZS_Q(A/D)=Q(1). Third, if A∩D=∅ then ZS_Q(A/D)=Q(0). Because of this, the method ZS obtain the expected values for the examples 3 and 4. Moreover this method obtain intuitively good results when dealing with non-coherent quantifiers and it is not an strict method. Finally, its computational complexity is O(nlogn), and the evaluation is faster and easier than Y_Q because Yager's method must obtain the values S_i and w_i previously.
It is clear that when D=X a type II sentence turns to be a type I sentence. We show in [4] that in this case ZS_Q(A/X)=ZS_Q(A).

We have also obtained a generalization of the probabilistic method GD to the case of type II sentences. In [5] we introduce GD_Q(A/D) to be

$$GD_Q(A/D) = \sum_{\alpha_i \in M(A/D)} (\alpha_i - \alpha_{i+1}) * Q\left(\frac{|(A\cap D)_{\alpha_i}|}{|D_{\alpha_i}|}\right)$$

where M(A/D) defined before is described as

$$M(A/D) = \{\alpha_i\} \qquad 1 = \alpha_1 < \alpha_2 < \ldots < \alpha_m < \alpha_{m+1} = 0$$

This method verify the same properties as method ZS, and hence it gives the expected results for the examples 3 and 4. The computational efficiency of this method is O(nlogn). Also we have that when D=X then GD_Q(A/X)=GD_Q(A).

We have generalized the methods S_Q and C_Q to the case of any quantifier, either coherent or not, and to the case of type II sentences. In [2] Bosc and Lietard compare both methods and conclude that the maximum difference between the results is 0.25. We have also shown that in the case of coherent quantifiers, these methods verify the properties we require for a good method. So S_Q and C_Q are both acceptable. Nevertheless there are more deep differences between their generalizations, GD_Q and ZS_Q , at least in the case of type II sentences, as the following example shows.

Example 5. Let us consider A={1/x_1, 0.01/x_2} and D={1/x_1, 0.99/x_2} and Q=∀. We have that x_2 is very close to be in D, but it is very far from being in A, so intuitively the percentage of elements of D that are in A should not be 1. We obtain a value coherent with that intuition from method GD, $GD_{\forall}$(A/D)=0.02. But with ZS we obtain $ZS_{\forall}$(A/D)=1. As we can see, the difference in the case of type II sentences is very high. In fact, we can make this difference as close to 1 as desired by decreasing A(x_2) and increasing D(x_2) without reaching 0 and 1 respectively.

Example 5 gives us a motivation to prefer the method GD to the method ZS. We think that ZS gives a counterintuitive result in this example because the functions maximum and minimum ignore most of the information contained in A and D, while method GD

takes into account all the information of A and D. This case also arises in the case of S_Q and C_Q, but the consequences are smoothed because the information about the referential X is fully used by both methods.

4. Conclusions

We have briefly reviewed previously existing evaluation methods for type I and type II sentences. We have discussed these methods from the point of view of some intuitive properties that we think any method must verify. We show by means of some examples that some of the previous methods are not appropriate in some cases, such as the case of sentences with non-coherent quantifiers. Zadeh's method is shownnto be very strict for the evaluation of both type I and type II sentences. We have developed generalizations of the best (to our knowledge) existing methods based on the Sugeno and the Choquet fuzzy integrals. These generalizations have better properties than previous ones. Finally, we have shown with example 5 that the method GD should be preferred in general because it takes into account all the information involved and (to our experiments) the results are always coherent with our intuition.

References

[1] Bosc, P. Lietard, L. Monotonic Quantified Statements and Fuzzy Integrals. *Proc.NAFIPS/IFIS/NASA Conference*, 8-12, 1994.

[2] Bosc, P. Lietard, L. On the Comparison of the Sugeno and the Choquet Fuzzy Integrals for the Evaluation of Quantified Statements. *Proc. of EUFIT '95*, 709-716, 1995.

[3] Delgado, M. Sánchez, D. Vila, M.A. Un Enfoque Lógico para Calcular el Grado de Cumplimiento de Sentencias con Cuantificadores Linguísticos. *Proceedings of Estylf'97*, Tarragona, Spain, 15-20, 1997.

[4] Delgado, M. Sánchez, D. Vila, M.A. Un Método para la Evaluación de Sentencias con Cuantificadores Linguísticos. *Proceedings of Estylf'98*, Pamplona, Spain, 193-198, 1998.

[5] Delgado, M. Sánchez, D. Vila, M.A. Fuzzy Cardinality Based Evaluation of Quantified Sentences. Submitted to Int. Journal of Approximate Reasoning.

[6] Delgado, M. Sánchez, D. Vila, M.A. A Logic Based Definition of Fuzzy Cardinality. Submitted to Fuzzy Sets and Systems.

[7] Liu, Y. Kerre, E. An Overview of Fuzzy Quantifiers. (I). Interpretations. *Fuzzy Sets and Systems* 95, 1-21, 1998.

[8] Liu, Y. Kerre, E. An Overview of Fuzzy Quantifiers. (II). Reasoning and Applications. *Fuzzy Sets and Systems* 95, 135-146, 1998.

[9] Yager, R.R. On Ordered Weighted Averaging Aggregation Operators in Multicriteria DecisionMaking. *IEEE Transactions on Systems, Man and Cybernetics* 18 (1), 183-190, 1988

[10] Yager, R.R. Fuzzy Quotient Operators for Fuzzy Relational Databases. *Proc. of IFES '91*, 289-296, 1991

[11] Zadeh, L.A. A Computational Approach to Fuzzy Quantifiers in Natural Languages. *Computing and Mathematics with Applications* 9 (1), 149-184, 1983.

EVALUATING RELIABILITY AND RELEVANCE FOR WOWA AGGREGATION OF SLEEP APNEA CASE DATA

David Nettleton
Computer Languages &
Systems Department
University Polytechnic of Catalunya,
Spain
netleton@lsi.upc.es

Lourdes Hernández
Respiratory Disease Institute,
Hospital Clinic
University of Barcelona,
Spain
fburgos@medicina.ub.es

Summary

In this article, joint medical and data analysis expertise is brought to bear using contrasting data analysis methods and the WOWA aggregation operator to solve a difficult medical diagnosis problem, that of sleep apnea syndrome screening. We describe a method of calculating the relevance and reliability weights used by the WOWA operator.
Key Words: sleep apnea diagnosis, questionnaire responses, WOWA aggregation, clustering, classification, reliability, relevance.

1. INTRODUCTION

Screening of Apnea cases is a difficult diagnosis problem, at present not satisfactorily resolved by standard statistical modelling techniques. We propose that part of the problem is due to the inherent fuzzy nature of a significant part of the data: questionnaire responses. We use diverse clustering and classification techniques to establish the relevance and reliability of each variable, which is then given to the WOWA aggregation operator to generate an aggregated value for each patient with high correlation to the apnea diagnosis.

The article is structured as follows: in section 2, a brief clinical description of the sleep apnea syndrome is given; in section 3, the method of diagnosis using WOWA operators and the problem of assigning the relevance and reliability weights is detailed; in section 4, unsupervised clustering techniques identify variables used to partition the data; in section 5, supervised modelling techniques are used to generate a ranking of variables in order of significance to the diagnosis output; in section 6, WOWA aggregation results are detailed using the reliability and relevance weights derived in sections 4 and 5; finally, section 7 discusses some conclusions of this method and the applicability to apnea diagnosis.

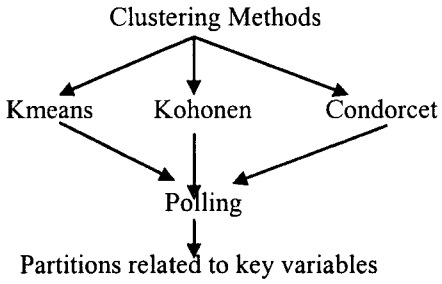


Figure 1a: Clustering Techniques determine relation of key variables to clusters

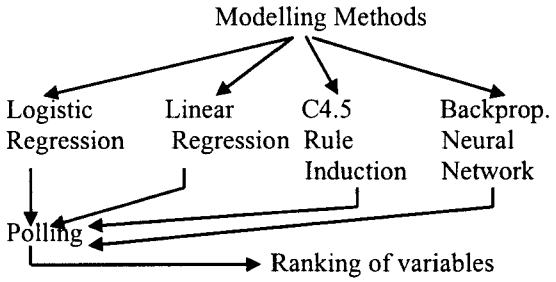


Figure 1b: Contrasting methods are polled to determine a ranking of relevance and reliability of the variables with respect to apnea diagnosis

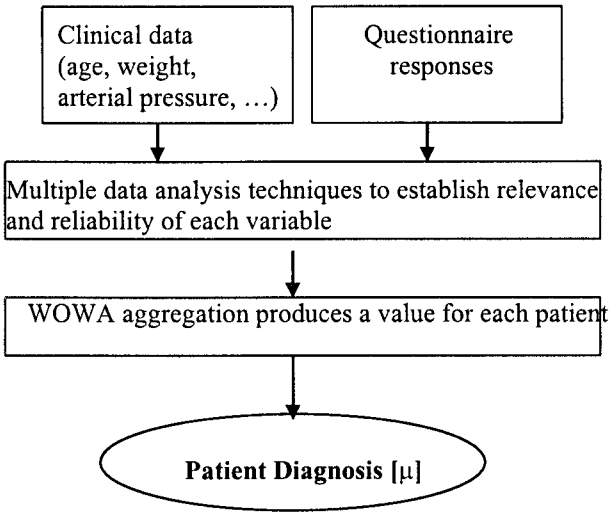


Figure 1c: Data processing of the apnea data input variables to produce a diagnosis

2. THE SLEEP APNEA SYNDROME

The Obstructive Sleep Apnea Syndrome (OSAS) is a set of secondary clinical manifestations relating to the ceasing (apnea) or reduction (hypopnea) of air flow during sleep, caused by a partial or total collapse of the upper air way at the faringe level. The severity of the OSAS is defined by the *apnea hypopnea index* (AHI) or the number of apneas plus the number of hypopneas per hour during sleep. Generally an AHI ∃10-15 is considered pathological.

2.1 DIAGNOSIS

The predictive value of the clinical data in OSAS diagnosis is low. Hoffstein [1] published results that indicated that clinical data explains 36% of the variability of the IAH (apnea hypopnea) and Katz [3] reported a figure of 39%, other authors report lower figures. The subjective clinical evaluation of the interviewer has also been evaluated and tends to have a low sensibility and specificity, in the order of 55%-65% respectively, for correctly classifying the sick. On the other hand, The predictive models for IAH based in clinical data have a higher sensibility of up to 90%. Their specificity, in the best of cases, does not reach 70%. At present, it is not appropriate to define rigid diagnostic criteria in this rapidly developing area. Neither is it possible to identify the ideal equipment for sleep studies.

2.2 APNEA PATIENT DATA

We have the collected data of the standard sleep patient questionnaire, for 154 patients, captured over a 1 year period. The data set contains 68,2% positive outcomes and 31.8% negative outcomes. The questionnaire consists of two main sections: the first records clinical data (age, weight, blood pressure, etc. ..); the second section consists of 41 questions to which the patient responds. The questions are divided in 3 subsections: 15 general sleep questions {s1...s15}, 16 respiratory related questions {r1...r16} and 9 somnolence related questions{s1...s9}. Based on this information, the doctor then gives a clinical evaluation: healthy; simple snorer; doubtful; typical apnea; other illness. We simplify this to: typical apnea; no apnea.

3. ESTABLISHING RELIABILITY AND RELEVANCE FOR WOWA AGGREGATION

We consider applying diverse data analysis techniques to the variables which have been collected for apnea patients. We wish to establish, for each variable, its relevance with respect to the apnea diagnosis, and its reliability. The first value will be the ρ weights in the first WOWA vector, the same vector being used for all variables and cases. The second value is represented by one ω vector per variable. Also we wish to establish the most significant variables with respect to apnea diagnosis. The four classification techniques used are: C4.5 rule induction, backpropagation neural network, logistical and linear regression. The techniques have been chosen so as to contrast the results given by significantly different approaches (Figures 1a, 1b, 1c).

3.1 WOWA AGGREGATION

In [5], Nettleton evaluated different aggregation methods for diagnosing sleep apnea. The following aggregation methods were considered: Ordered Weighted Average (OWA) [7], Weighted Ordered Weighted Average (WOWA) [6] and Principal Components. OWA uses a vector in which, for each variable, a value is assigned which indicates its relevance. In WOWA a second vector is incorporated whose values indicate the reliability of each variable. In this article we extend the previous work, calculating the weights from statistical analysis of the data.

3.2 RELEVANCE AND RELIABILITY

Relevance is a standard data analysis objective for which we can apply diverse algorithms and interpret the results. Relevance is more straightforward to establish in statistical data analysis, than reliability.

Reliability is influenced by different aspects. There are data aspects, such as %missing and %erroneous. Then there are application dependant aspects which, in the case of the questionnaire responses can be the truthfulness with which the patient responds (it may be that if the patient goes to sleep at the wheel of a car, s/he does not wish to make that known, and thus tends to give a higher negative response rate to this question {**s5**}, than it really should have). Each variable has a ω vector which, on ordered values, heightens some values while dampening others. We execute the WOWA aggregation function with an Ai

data vector for each case, the Vρ vector remains constant and represents the *relevance* of the variables, while the $V\omega_n$ vector represents the *reliability* of the ordered values of each variable and varies for each variable. For example, we call WOWA with (A_i, Vρ, $V\omega_n$), where A_i is data row **i**. Also, $V\omega_1$=even bias (E); $V\omega_2$=bias on low values (L); $V\omega_3$=bias on high values (H); $V\omega_4$=bias on medium values (M); $V\omega_5$=bias on high & low values (O). For example, $V\omega_1$=even bias (E) could be assigned as {0.2, 0.2, 0.2, 0.2, 0.2} for the five data values belonging to variable 1, while $V\omega_3$=bias on high values (H) could be assigned as{0.1,0.1,0.2,0.3,0.3}, for variable 3.

4. UNSUPERVISED CLUSTERING AND STATISTICAL TECHNIQUES

With reference to Table 1 (below) we can see that in methods 1, 2 and 3, *partner* has influenced partitioning. Kohonen y Kmeans tend to have biased the *'g'* responses while Condorcet has used the *'s'* responses more for partitioning. There is not a clear consensus between the different clustering and statistical techniques. Methods 1, 2 and 4 used numerical representation of all the data, while method 3 used a categorical representation with ChiSquared for the significance tests. **Kohonen net:** Various architectures of net were tried: input layers of 41 neurons (questionnaire responses only), 27 (clinical data only), and 68 (questionnaire responses and clinical data). **Kmeans:** Standard SPSS Kmeans was used for 2 clusters, maximum iterations set to 100, convergence at 0.02. **Condorcet – mixed data type clustering:** A proprietary IBM algorithm based on the *Condorcet* [2] distance criteria was used to generate 9 segments. All data was prepared as categorical and a chisquared measure was used to rank the variables in each segment and between segments. **Cross product covariances (Pearson):** A standard SPSS numeric covariance was used with the pearson product moment option, to generate covariances between all the variables, defined as numeric.

5. SUPERVISED CLASSIFICATION AND STATISTICAL MODELS

We contrast four techniques, each using a different algorithmic basis, with the objective of realising a consensus for the variables being evaluated. With reference to Table 2 (below) we can see that methods 1, 2 and 3 have identified *waist* as a significant attribute, while methods 1, 3 and 4 have identified *g1* as significant. Other identified variables are *r2*, *partner*, *weight* and *s10*. **C4.5 rule induction:** Quinlan's standard C4.5 algorithm was used, with 25% pruning, no external test set, and no grouping. **Backpropagation neural network:** The neural network training phase generates a *sensitivity analysis* which provides a ranking of the variables with respect to the output (in this case, the diagnosis yes/no). **Logistic regression:** Standard SPSS logistic regression was used with 3 test models. Overall precision's were: 89,66%, 88% and 75%. **Linear regression:** One SPSS linear regression was generated. The R^2 value was 0.31309, the standard error was 0.51035.

Table 1. Clustering and statistical techniques applied to the apnea cases and the identification of key variables which distinguish the resulting partitions

	Kohonen (1)	Kmeans (2)	Condorcet (3)	Cross product covariances (Pearson) (4)
Most significant variables	partner, weight, g1, r1, g4, s5 (2 and 6 clusters)	partner, sex, g4, r6, g13, g5, s5 (2 clusters)	s5, s2, s1, s6, r13, partner, g6, g7 (9 clusters)	neck, weight, age, alcohol

Table 2. Significance ranking of input variables for different methods

	Logistic regression (1)	Linear regression (2)	C4.5 rule induction (3)	Back propagation neural network (4)
Most significant variables	neck, g1, partner, s9, s8, s7, s6, s10, waist, r12, r2, r5, r6, g2, g6	g8, partner, waist, hip, weight	r3, r2, waist, age, weight, g1	sex, r15, g10, g1, r9, r1,hta,tabacco, height, alcohol, weight, r3, r8, s7, g5, s10, r2, r5

6. WOWA AGGREGATION

Nettleton, in [4] defined, jointly with the medical expert, a most significant subset of 15 variables for apnea diagnosis. This was: {age(0.5, E), sex(0.7, E), weight(0.7, M), imc(0.7, M), neck circumference(1.0, E), alcohol(0.5, M), hta(0.7, E), r1(0.9, H), r2(0.9, H), r11(0.9, M), r13(0.9, M), s3(1.0, M), s4(1.0, M), s5(1.0, M), s6(1.0, M)} with figures in brackets being the respective relevance and reliability weighting profiles, also assigned by the medical expert. The reliability weights are coded as indicated in Section 3.2. Thus each variable has its own characteristic vector of reliability weights which emphasise the more credible responses to a question and dampen the less credible/certain ones. For example, a response of 'always', (assigned in the data the numeric value '5') to question **s6** may be unusual and therefore we would lower its reliability rating, by giving it a smaller proportion of the ω weight. The above reliability and relevance values have to be prepared for input to WOWA so that they sum to 1 for each case. The values are reduced proportionately to achieve this. The normalised value is converted into the relevance and reliability weights for WOWA which respectively sum to 1 for all variables, that is , $\Sigma\rho = 1$, $\Sigma\omega = 1$, where ρ is relevance and ω is reliability.

With reference to Table 3 (below) we can see a favourable result for diagnosis of positive cases and a good result for negative cases, in comparison with the methods used in the literature[1][3].

Table 3. Correlation of WOWA with Apnea Diagnosis for three different weight assignment methods for reliability and relevance

Weight assignment method	Expert assignment of weights	Data analysis assignment of weights	Expert + data analysis assignment
Diagnosis of positive cases	0,75	0,78	0,81
Diagnosis of negative cases	0,65	0,61	0,67

7. DISCUSSION OF RESULTS AND CONCLUSIONS

The resulting consensus from all data analysis methods (tables 1 and 2) indicated the following 9 most significant variables and their corresponding relevance and reliability weighting profiles: {partner(0.70, E), weight(0.70, M), neck(0.92, E), g1(0.68, M), s5(0.95, M), sex(0.7, E), r15(0.60, M), hta(0.67, E), r5(0.90, M)}. Question **g1** is: *"how many hours do you normally sleep?"*; question **s5** is *"do you fall asleep while driving on the motorway?"*; question **r15** is *"do you have lapses of memory or loss of concentration"* and question **r5** is *"have you noticed an increase in the intensity of your snoring recently?"*. Executing the WOWA aggregation with the above input weighting vectors and the 154 patients case data rows, produced the results of table 3. In table 3 the output aggregated value produced by WOWA has been correlated with the binary value which represents the apnea diagnosis.

In this article we have described a method for establishing the relevance and reliability weights needed by the WOWA aggregation operator, applying them to complex data of a real medical problem. With the WOWA aggregation method to apnea diagnosis we can include relevance and reliability information in a more precise manner, to improve the success rate for correct diagnosis.

The authors are grateful to the Medical Faculty of the University of Barcelona, and to Vicenc Torra of the Institute for Investigation in Artificial Intelligence, Bellaterra, Spain, for their collaboration.

References

[1] Hoffstein, V., Szalai J.P. *"Predictive value of clinical features in diagnosing obstructive sleep apnea"*. Sleep 1993; 16: 118-122.

[2] IBM Data Management Solutions White Paper. IBM Corp., 1996.

[3] Katz I., Stradling J., Slutsky A.S., et al. *"Do patients with sleep apnea have thick necks?"* American Review of Respiratory Diseases, 1990; 141: 1228-1231.

[4] Nettleton, D.F., *"Attribute fusion using a heterogeneous representation of crisp and fuzzy medical data"*. Eighth International Fuzzy Systems Association World Congress. Taipei, Taiwan, 1999.

[5] Nettleton, D.F., L. Hernandez. *"Improving questionnaire screening of sleep apnea cases using fuzzy knowledge representation and aggregation techniques"*. Research Report Ref. LSI-99-29-R, Dept. Computer Languages and Systems, University Polytechnic of Catalunya, Spain (1999).

[6] Torra, V., *"The Weighted OWA Operator"*. International Journal of Intelligent Systems, Vol. 12, 153-166. John Wiley & Sons (1997)

[7] Yager, Ronald R. *"Families of OWA operators"*. Fuzzy Sets and Systems 59 (1993) pp125-148.

IMPLEMENTING A FUZZY PROCESSOR ON PROGRAMMABLE LOGIC CIRCUITS: MODULARIZATION CRITERIA

Esperanza M. López
E.T.S. de Ingenieros Industriales
Universidad de Murcia
Cartagena. Murcia
espe@plc.um.es

Alfonso Aniorte Carbonell
E.T.S. de Ingenieros Industriales
Universidad de Murcia
Cartagena. Murcia
aac@plc.um.es

Ramón Ruiz Merino
E.T.S. de Ingenieros Industriales
Universidad de Murcia
Cartagena. Murcia
ramon@plc.um.es

Summary

In this communication, we propose a scheme for implementing a fuzzy processor on programmable logic circuits (FPGA), utilizing pipelined structures. We have partitioned the processor structure in several sections, which correspond to different integrated circuits. Such I.C. carry out the operations inherents in a fuzzy process (fuzzifiers, inference engine, defuzzifiers) in a very efficient an flexible manner maintaining a low cost per prototype.

Keywords: FPGA, Fuzzy Processor, ASIC, Digital Design.

1 INTRODUCTION

The need of making compatible the calculation requeriments that Fuzzy Logic imposes, with the limitations in response time of certain applications, in particular the control of processes, implies almost unavoidably to find alternatives to the conventional calculation systems, which are not quite adequate for the efficient treatment of the operative schemes that arise of the linguistic variables type. Fortunately, the modularity and the simplicity of the fuzzy processing algorithms which are adequate for these applications, permit their efficient implementation on dedicated hardware structures [2]. The problem is usually to decide how much flexibility (measured in terms of the [programmability] of the different blocks that constitute a fuzzy system) should be supported in a dedicated fuzzy device. Factors like the speed, complexity, restrictions of response time and interaction requirements with not fuzzy algorithms, will influence finally on the decision [1].

The implementation of digital systems based on regular structures of gates or, more in general, of cells, has revolutionized some conventions of the digital logic and arithmetic. Viewed as specific calculation units, their implementations provide excellent performances in the majority of the applications, providing true generic hardware.

Within this methodological alternative for implementing of electronic digital systems, the introduction by the middle of the 80's of the Field Programmable Gate Array (FPGA) meant a milestone in the field of the design of application specific integrated circuits (ASIC). In fact, at the present time their technology makes them competitive in both complexity and versatility as opposed to other ASIC alternatives for direct implementation on silicon (standard cells or gate array). On the other hand, these circuits present clear advantages in terms of: time of development, flexibility, and specially, cost of the prototypes. To sum up, even though a full custom implementation on silicon usually surpasses the capacity of a generic programmable system and reaches higher speed, very few applications deserve the cost of creating a specific integrated circuit, being more and more attractive the solutions based on FPGA.

We present in this work the general structure of a configurable fuzzy processor, developed to be implemented on several FPGA circuits. Given the flexible nature of this implementation style, the proposal that we make does not constitute a closed solution; rather it pretends to open perspectives in both the modularization of the design and the possibilities of pipelining of the operation, starting from a library of constructive blocks which allow to construct multi-FPGA, specific processors adapted to the dimensions and particularities of the fuzzy computation problems. We have intended to evaluate the possibilities of this methodology of implementation, in a field so demanding and mature as that of the dedicated circuits in fuzzy logic. It will be necessary to take advantage of the characteristics of flexibility, modularity and parallelism of the modern programmable devices in order to achieve competitive implementation features for speed and complexity

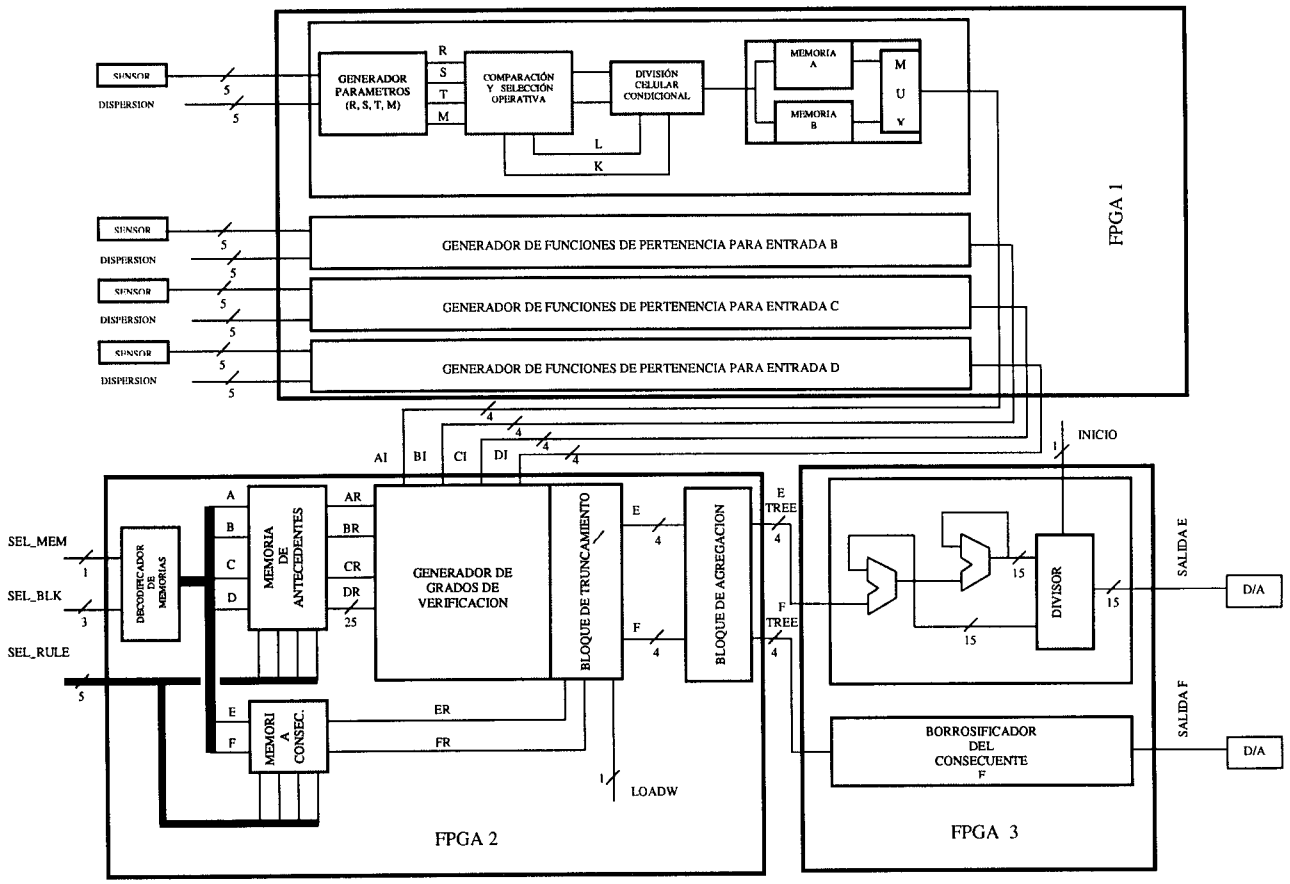


Figure 1: General block diagram of the fuzzy processor.

2 MODULAR STRUCTURE OF PROCESSOR

The processor adjusts to the classical structure, in which parallelism is introduced by means of a pipeline of five stages. The main elements of the architecture are summarized in the general block diagram shown on fig. 1. In this figure we can distinguish three main modules, that correspond to the physical blocks in which we have divided the processor. This modularizationn of the design in physical blocks differentiated in three levels, together with the characteristics of configurability introduced in certain sections of the design, confer the proposed architecture a high grade of flexibility, allowing several options with regard to the number of inputs, number of rules and number of outputs. In this way, the block diagram of the fig. 1 (one integrated circuit by level) only corresponds to the simplest configuration, adequate for systems of rules consisting of 4 antecedents (or 2 in the alternating configuration of structure of rules) and 2 consequents (or 1), containing 25 rules (or 50 when we process rules 2 to 1).

2.1 MODULE OF INPUT GENERATION

This module carries out the translation of the variables measures from of the sensors in possibility distributions consisting of 32 values quantified in 4 bits; such possibilities are the inputs to the inference process. The characteristics of these triangular distributions are determined by both the measurement of the sensor and the value of dispersion that is introduced in order to characterize the uncertainty in the measurement of each input. Both values are quantified in 5 bits.

In order to produce the distributions, we have designed a generator of possibility values that avoids the demands of memory of the systems based on Look up Tables [3], that could constitute a problem when implemented on FPGA, given its limitations as for available RAM. The generation of the membership values is carried out starting from the calculation of four parameters (fig. 2),which are introduced in a module of comparison and selection which determines the function to be applied in each case. In the intervals where it is necessary to carry out a division (to calculate the

lineal interpolation), this is performed by means of a cellular unit whose characteristics of regularity and local connectivity favor their implementation, providing some excellent speed capabilities. In the cases of absence of dispersion in the measurement, the need for these operations of division will are eliminated.

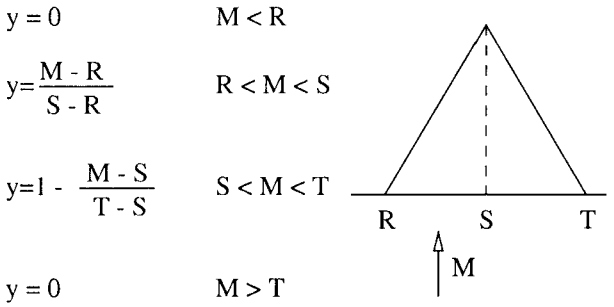


Figure 2: Parameters of the membership values.

The possibility values obtained in this way are stored in one of the two halves of a memory area, from which they will be introduced in a later phase into the module of rules to carry out the fuzzy inference. The inclusion of a double memory structure, where reading and writing of data are alternatively switched, allows us to introduce a first pipeline stage in the controller operation. So, as long as the A half receives the data of the generator in a certain time interval (writing), simultaneously the B half will supply the possibility values (reading) of the previous measurement to the inference.

The moderate complexity of this module allows their implementation in a FPGA of medium size, of the type XC4010 (10000 logical gates and 12800 bits of RAM). This circuit will be occupied about 90%. Concerning to the frequency, the maximum delay in this module corresponds to the structure of division (170 ns). This delay determines the maximum operation frequency of the system.

2.2 MODULE OF RULES (INFERENCES)

In this module the distributions resulting from the input evaluation through the set of rules are calculated. Such rules can be configured as 4 antecedents and 2 consequents rules, or alternatively as 2 antecedents and 1 consequent rules; in the first case to a maximum of 25 can be stored, and in the second 50 ones. The rules are stored in a RAM memory, value by value, so that the total dedicated area to this purpouse is 25 x (4+2) x 32 4-bit words (4800 x 4).

The calculation of the verification degrees among the inputs and the antecedents of the rules is carried out by means of a direct implementation of the Mamdani's algorithms; inputs are compared in parallel to all the antecedents of each rule. The minima of the membership functions are extracted sequentially element by element and then the maxima of this series are stored in registers. In a later stage of the pipeline these degrees of verification will truncate the consequent distributions, passing these truncated distributions to an aggregation tree (MAX), that will produce the output distribution associated with each consequent. The aggregation is made up as a strictly combinational function, and it contains the critic path which determines the maximum operation frequency of the device. This frequency is higher than the one imposed by the module of input generation. The distributions resulting from the aggregation block constitute the input to the following module.

Such a method for storing rules permits a high degree of freedom in the definition of antecedents and consequents distributions. Concerning to the number and structure of the rules, we also have a high versatility; the possibility of chaining rules could be effective just applying the output from one of these modules to the input of circuits a second level. Concerning to the addition of in rules, the disposition of circuits in parallel connected to the same module of generation of inputs will proportion this possibility, being necessary a minimum amount of logic (MAX modules) in order to complete the aggregation tree; such logic can be included within any of the FPGA of this level, as their degree of occupation is not too high. In order to implement these modules, the programmable devices we must use have to be quite complex. Devices of the type XC4052 permit this implementation, being occupied about the 75%. The XC4052 has about 52000 logical gates, and an available RAM of almost 62000 bits. On the other hand, the maximum operation frequency of this module is higher than the one determined by the module of input generation (about the double). So the frequency of the whole system is not limited by this module.

2.3 DEFUZZIFIER MODULE

The last module contains a double defuzzifier structure that simultaneously processes the two output from the module of rules. For the defuzzification, we use the center of gravity method, what makes necessary a new operation of division. In this module, the division has been implemented according to a sequential algorithm without restoration. The use of a division structure different from that of the module of rules is due to the fact that the operative scheme utilized for the calculation of the center of gravity permits to synchronize the output rate of the module of rules with the displacement/accumulation steps involved in this division al-

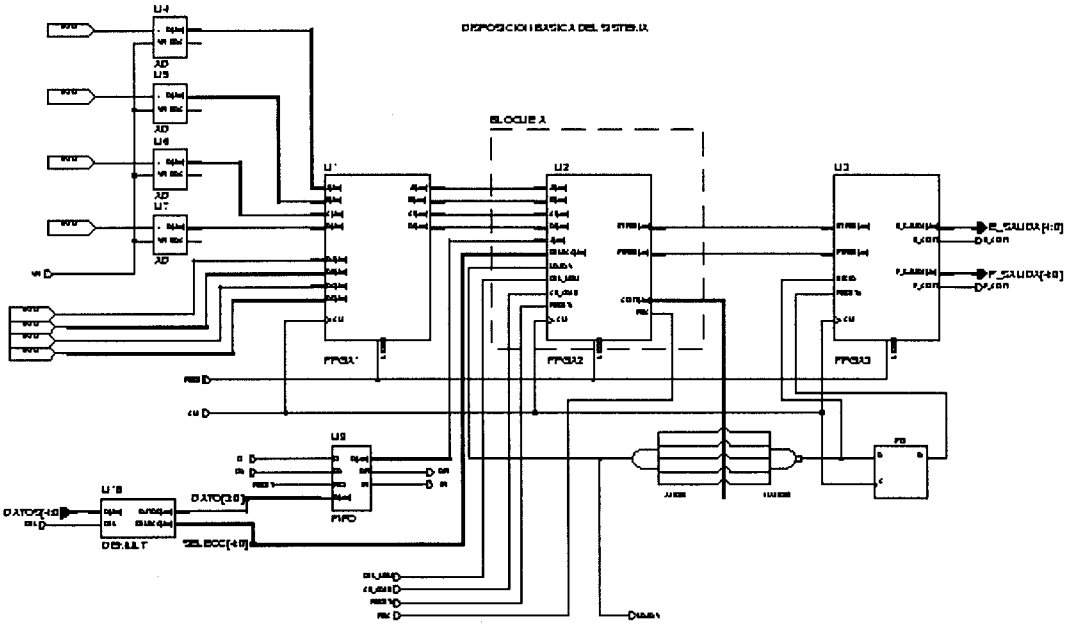


Figure 3: General structure of the controller card.

gorithm. Each of such steps involves a propagation delay much lower than the one of the cellular divisor included in the first module (delay not exceeds 40 ns in this case). This third module doesn't offer problems of complexity, so we can opt again to utilize a FPGA of medium size (type XC4010), that allows to hold the functions of the two defuzzifiers easily.

3 CONTROLLER CARD

The three integrated circuits we have presented constitute a versatile fuzzy processor as a whole, providing and adequate platform to carry out control tasks based on fuzzy rules. However, for this operation it is necessary a certain logic that implements the communication with a ghost computer in order to program the integrated circuits. Complementary logic elements are also needed to configure the integrated circuits and to provide an adequate sequencing of the operations. The fig. 3 represents the disposition of these integrated circuits in a controller card and their connection. The proposal includes the programming of the integrated circuits through a parallel port, that will also serve for storing the rules in the memory of the second integrated circuit.

4 RESULTS AND CONCLUSIONS

The integrated circuits that implement our proposal have been verified using a temporal analyzer, in order to determine the critical path and, consequently, the maximum frequency to which they can operate. The results of this analysis have been commented previously, being the first of the modules who imposes a greater limitation in the operation frequency. A functional analysis has also been carried out. Problems of temporal violation detected in first instance have been solved, and we have verified the correctness of each of the modules as well as and the fuzzy processor as a whole. In summary, we consider satisfactory the results obtained, confirming that the elected approach is a suitable one, competitive in the majority of aspects with ASIC solution of much greater cost per prototype.

References

[1] A. Costa y A. DeGloria (1995). Hardware solutions for Fuzzy Control. *Proc. IEEE* pp 83.

[2] A. Gabrielli et al.(1997). VLSI Design and Realization of a 4 Input High Speed Fuzzy Processor. *6th IEEE Int. Conf. Fuzzy Systems.* Barcelona. July.

[3] H. Watanabe et al.(1990). A VLSI Fuzzy Logic Controller with Reconfigurable, Cascadable, Architecture.*IEEE Journal Solid State Circuits.* n.25.

USING FUZZY LOGIC TO IMPLEMENT ADAPTABILITY IN WAY-Z39.50

Camino Fernández **Vicente Matellán** **Ignacio Aedo** **Paloma Díaz**
Departamento de Informática
Universidad Carlos III de Madrid
Av. Universidad 30, E-28911 Leganés (Spain)
E-mail: {camino@inf,vmo@gsyc.inf,aedo@ia,pdp@inf}.uc3m.es

Summary

The design and implementation of adaptive systems for accessing information is a sophisticated task. Different models have been proposed to build systems that adapt to the user reducing the complexity in the information browsing process, saving user time and offering personalized interfaces. The work described in this paper, named WAY-Z39.50, is an example of fuzzy adaptive system specially built for accessing digital libraries.

Keywords: User interface, Agent, Adaptive system, Z39.50, Digital libraries, Fuzzy application, WAY.

1 INTRODUCTION

Finding and accessing information are complex problems, due to the inherent difficulty about locating appropriate information servers and the information itself; and to the complexity about presenting the information to the user through an adequate interface [1, 5].

To help in the design of this kind of systems, WAY [2] proposes a general model of architecture whose main goal is to improve information access systems through the fulfillment of three tasks: adaptation of the interface to the user, help in the searching process, and extraction and maintenance of updated data about the available information servers.

One implementation of the WAY general model is WAY-Z39.50. It is designed to access servers supporting the Z39.50 information retrieval protocol standard [4], that is being installed in most libraries all over the world to provide access to their catalogues. WAY-Z39.50 is based on an agent-decomposition method where each task of the system is managed by a different agent. Figure 1 shows the five agents that comprise the system, representing by means of dashed ellipses those agents that are instantiated for each user, and by solid ellipses those that are instantiated only once for the whole system.

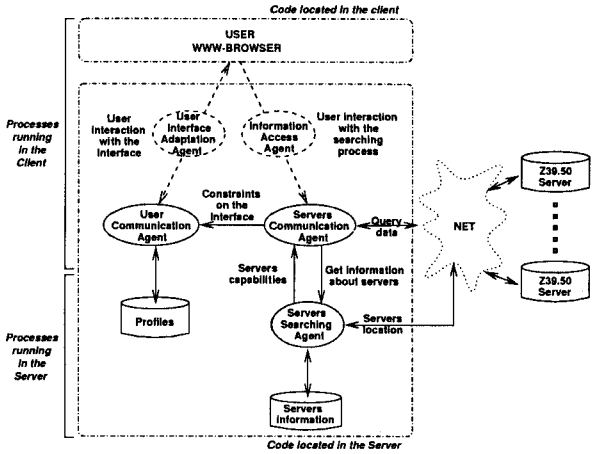


Figure 1: WAY-Z39.50 architecture

Three of theses agents implement a fuzzy [7] adaptive process (the first three ones in the following description), while the other two ones use crisp processing algorithms:

User Interface Adaptation Agent: The task of this agent is to spy the user actions to decide the best way to adapt the interface to the user. The agent takes decisions and transmits its suggestions to the user communication agent in order to perform them (building the interface). The user can always decide not to allow the agent to change anything.

Information Access Agent: The task of this agent is to spy the user searching process. When a user makes a query to one information server,

this agent looks for the servers she prefers and launches parallel searching processes to the ones considered more appropriate by the agent.

Servers Searching Agent: The task of this agent is to maintain updated information about the accessible servers. To perform this task, it can use sources such as web pages, mailing lists or USENET news groups. The agent periodically consults these sources and updates the servers' list. The period is also updated by the agent itself.

User Communication Agent: It is in charge of building the user interface according to user preferences, as well as transmitting the user's actions to the user interface agent following the user interface adaptation agent suggestions.

Servers Communication Agent: It communicates with the Z39.50 information servers, translating user inputs into adequate queries and passing the information retrieved to the user communication agent.

2 WAY-Z39.50 design and implementation

One of the requirements for WAY-Z39.50 implementation was that the software being developed should be broadly available. In order to fulfill this requirement, Java programming language was chosen to make WAY-Z39.50 available through a web page. A java applet implements the interface of the application that is built by the user interface adaptation agent. This agent communicates with the information access agent and the servers searching agent using Java-RMI (Remote Method Invocation) technology to support distribution.

The general model does not force concrete systems to use one particular mechanism to implement the different kinds of adaptability. In WAY-Z39.50, the mechanism for deciding which modifications in the user interface are going to be suggested is based on fuzzy decision systems. Suggestions about user interface comprise both the layout of the components, and their appearance. Besides, fuzzy decision systems will also be used to help the user in the information searching process, as well as in the classification of the Z39.50 servers available.

Fuzzy Relational Algorithms [6] (FRAs) are the method used to store the knowledge required to take the decisions (for instance, which server is going to be used). FRAs are composed by a finite set of fuzzy conditional statements (antecedent and consequent). Antecedent of conditional statements are conjunctions of fuzzy variables defined over inputs. For instance:

IF *Time_Server_1* **is** *Long* **and** *Errors_Server_1* **is** *High* **THEN** *Accessibility_Server_1* **is** *Bad*

The operation named *fuzzyfication* transforms a non-fuzzy quantity into a fuzzy variable. This operation is made using the definition of those variables, that in WAY-Z39.50 has been done using trapeziums [8]. For instance, a value as $Time_server_1 = 10msec$ will be translated into activation levels of fuzzy labels as *Fast*, *Slow*, etc.

An extension of the *modus ponens* [3] is used as the inference mechanism to assign values to the fuzzy output variables. These values can be translated into non-fuzzy quantities using a *defuzzyfication* process. The adopted defuzzyfication operation is a version of the "Center of Gravity" procedure [8].

2.1 Designing the fuzzy relational algorithms

Several automatic methods such as neural networks or genetic algorithms have been proposed to generate FRA in different domains. The heuristic method has been the one used in the design of WAY-Z39.50. "Heuristic" means that a human designer will use her experience in order to define the fuzzy subsets (trapeziums in the current implementation) and to write the rules. The main reason for using this approach was to take advantage of the designers' experience.

Three different FRAs have been designed in the current implementation. Each one is used as the processing mechanism of each of the main WAY-Z39.50 agents.

First, the User Interface Adaptation Agent uses a FRA to adapt the user interface taking as inputs the detailed actions of the user during the sessions. The input of this FRA is any action performed by the user on the interface, that is, using one particular searching field, visiting first one kind of records or visiting only the records written in one language. Its output also affects interface aspects as, for instance, placing the searching fields in a different way, ordering the records in a special manner or showing only the records written in the user mother tongue.

Second, the Information Access Agent uses the FRA to decide how many simultaneous searches in different servers are going to be made to speed up the searching process. The agent has to take into consideration, as inputs of this decision system, issues such as the interest of the user for previously proposed searches in that server, the current load of the machine or the accessibility of the alternative servers.

Third, Servers Searching Agent uses a FRA to keep an ordered list of available Z39.50 servers taking into account aspects such as the speed of the communication with that server, its accessibility -as some servers are not available the whole day- or the quality of the results obtained by the user from that server.

2.2 Implementation issues

The software developed is implemented in Java 1.1.5 and it is structured into a hierarchy of packages. The root of this hierarchy is the `way` package, whose main elements are the subpackages `way.interfaz`, that implements the interface building facilities, `way.z3950`, that implements the client side of the Z39.50 protocol, and `way.agents`, that implements the agents of the system. The most relevant java subpackage, that concerns fuzzy issues in WAY-Z39.50 code, is `way.agents.reasoning` which implements the FRA. This implementation of FRA comprises the definition of fuzzy input variables, by means of their name and their membership functions (defined as the four points of a trapezium of height 1) defined in their range. The same format is used to describe fuzzy output variables and their labels.

Fuzzy `rules` are also represented as a separate java class made up by the `antecedent` and the `consecuent`. Each of them is composed by a set of `conditions` (represented by another class). These conditions correspond to membership relations. The `rule` class includes a method to implement evaluation which invokes the fuzzyfication of the components, then performs the modus ponens, and finally invokes the defuzzyfication process.

Given this implementation, the definition of FRA consists of the enumeration of the inputs, describing the labels using four points which define the trapezium; the enumeration of the outputs in the same way; and the specification of the fuzzy rules which describe how the outputs are modified depending on the inputs.

3 A session with WAY-Z39.50

When a user first enters the system, she is asked to enter a user name and a password. These data do not have a security purpose, since the system may be freely accessed, but they are only used to identify the user and the agents that are assigned to her (a user adaptation agent and an information access agent). These agents are in charge of the adaptation process and the storage of the user preferences that are used in the reasoning process.

At this point, the user interface adaptation agent builds a particular interface for that user. If this is the first time the user enters the system, a default interface is shown. This first interface depends on the kind of machine the user is connecting from and on the default language selected in the navigator used to access the system. During the session, the user interface adaptation agent is also active and spying user actions. The observation process generates information that is used as input for the fuzzy systems. For instance, the fuzzy reasoning process performed by the user interface adaptation agent can obtain through the fuzzy process a higher value for the `importance` variable of the field `title` than for the field `author`. If the field `author` is displayed in the first position and the field `title` in the second one, this agent would suggest to interchange their positions. Some of the inputs that affect the `importance` of a field are, for instance, the `use` of the field and its `success_rate`. The `use` of the field is the percentage of times that the field is used to make a query and the `success_rate` depends on variables such as `searching_success` that again depends on other variables as `response_time` or `long_records_visited`.

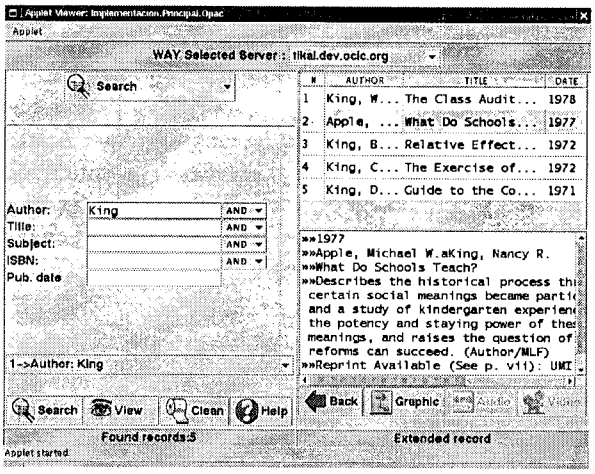


Figure 2: WAY-Z39.50 user interface

The user can always accept or refuse proposed changes, as some users do not want computers to take decisions that affect their work without their permission. Let's show a couple of examples about the possible adaptation. A user may have visited only books records so, as the `books_records_frequency` would have a very high value, the agent would propose the user to order the records beginning with the ones corresponding to books. If the same user consults mainly records in Spanish, the `spanish_records_frequency` would get a higher value than the English or French ones, so the agent would propose to order the records beginning with books in Spanish and following with books in English and so on. Figure 2 shows a WAY-Z39-50 customized interface where the user has made a query

and is visiting the extended records.

When the user makes a bibliographical query to the system, the first result obtained is the number of records satisfying the query. The user can consult a summary of the results as short records (upper right area of figure 2) meaning that she finds acceptable the number of records obtained. If the information shown in these records seems interesting, she can visit their extended version (bottom right area of figure 2). In this case, it can be supposed that she has found some of the information required or, at least, an interesting part of it, so variables as `server_usage` are increased.

While the user is querying the system, the information access agent is spying the process, which means that it is getting the values of the input variables to the fuzzy reasoning process. This way, the agent has information about the servers that the user visits more frequently or give her back the best results -where "best results" depends on the `server_delay` and `long_records_visited`. While the request is being sent to the Z39.50 server by our system, this agent is deciding and launching new searches on the servers considered more suitable for that user. Among others, variables as `server_accesses`, the number of times that the user has accessed a server, `server_success`, the percentage of successful connections to that server, `server_speed`, the speed of the connection, `long_records_visited`, the number of long records looked up in that server, etc. serve as input variables from which the `server_qualification` is obtained.

The searches launched by the agent are suggested to the user indicating the number of records found matching her query in other servers. This way, the user saves the time that she would have employed in making another query in the same server or going to another server and repeating the query, as it should be done in the classical systems that integrate many library servers.

Finally, the user can choose to go deeper into one of the proposed searches, refine the query or choose another server, beginning the whole process again. From any of these decisions, the agents will collect information that will be used as inputs for future operation as input variables to the fuzzy reasoning process that affect, among others, the `server_qualification`.

4 Conclusions

One of the main advantages of using WAY based systems, and particularly the advantages provided by WAY-Z39.50, is that their interface adapts to the user actions while she uses the system. To do that, the only information source are the user actions themselves, which means that there is no need to ask the user in any moment. They also offer alternative searches in the servers preferred by the user that respond best to information requests, saving user time.

The use of fuzzy logic in the adaptive process has made easier the task of specifying the rules to modify the user interface and to participate in the searching process. The logic rules are written in almost natural language, and the variables and labels involved in those rules have also got a well-known meaning, which makes them easy to design. These rules are also easily accessible as they can be stored on a text file or on a data base. A debugging process makes sure that the rules are correct and notifies the user if they are not. Finally, the reasoning process is really lighter than other mechanisms that require heavy computations, such as genetic algorithms or neural networks.

References

[1] D. Benyon and P. Palanque (1996). Critical Issues in User Interface System Engineering. *Springer-Verlag.*

[2] C. Fernández and P. Díaz and I. Aedo (1998). WAY: An Architecture for User Adapted Access to Z39.50 Servers based on Intelligent Agents. *Proceedings of the European Conference on Digital Libraries 98.* pp. 665-666.

[3] E. H. Mamdani (1984). An analysis of formal logics as inference mechanism in expert systems. *International Journal of Man-Machine Studies.* Vol. 21, n. 1, pp. 213-227.

[4] J. Michael and M. Hinnebusch (1995). *From A to Z39.50: A Networking Primer.* Mecklermedia.

[5] B. Shneiderman (1998). *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* Addison-Wesley Publishing Company, 3rd Edition.

[6] L. A. Zadeh (1973). Outline of a New Approach to the Analysis of Complex Systems and Decision-Making Processes. *Transactions on Systems, Man and Cybernetics.* Vol. 3, n. 1, pp 28-45.

[7] L. A. Zadeh (1988). Fuzzy Logic. *IEEE Computer.* Vol. 21, n. 4, pp. 83-93.

[8] Hans-Jurgen Zimmermann (1990). *Fuzzy Sets. Theory and its Application.* Kluwer Academic Publishers.

GENERATION AND EXTENSION OF MAPS OBTAINED BY AUTONOMOUS ROBOTS

Maite López-Sánchez, Ramon López de Màntaras and Carles Sierra
Institut d'Investigació en Intel.ligència Artificial, IIIA-CSIC
Campus UAB, 08193 Bellaterra,
{maite, mantaras, sierra}@iiia.csic.es

Abstract

We present some improvements of a previous environment mapping work. Firstly, we relax a previous restriction about the environment orthogonality. And secondly, we increase the environment coverage by extending occupancy information in the map. Finally, some results show how that extension process yields to safer paths when extended maps are used for planning.

Key words: autonomous robots, possibility theory, map generation.

1 INTRODUCTION

The background of this paper is a previous mapping work [5] which consisted on two different steps: first, a troupe of small autonomous robots explore an unknown office-like orthogonal environment, and second, a host computer receives exploration information from the returning robots and uses it to build a map. Robots explore moving randomly in clear areas and following walls (or obstacle edges) when detected by their infrared (IR) sensors. The computer host models the environment generating a grid map in terms of degrees of possibility and necessity [2] of the position of the detected walls and obstacles. Since a total coverage cannot be guaranteed, the host also computes the shortest path between two positions so that a robot can use it as a guide to reach a less explored area by combining the path information with reactive behaviour.

This paper describes two improvements of the mapping: one is that the assumption about the orthogonality of the environment is partially discarded and the other is that detected segments are enlarged on the basis of implicit map information. This map extension process increases the map coverage of the real environment and gives more conservative and informed paths that reduce the robots use of reactivity when following them.

2 NON-ORTHOGONAL ENVIRONMENTS

Our mapping process only considers environmental features with edges long enough to be followed by the robots. Small obstacles, such as chair or desk legs, are not represented. In fact, since they can be avoided using reactivity, they are not crucial.

Considering only relatively-large environmental features, we can assume that office-like environments are highly orthogonal. On the one hand, walls are usually connected by right angles. And on the other hand, human made objects in an office such as bookshelves or drawers tend to have rectangular shapes. In case a robot follows a wall or an edge of an obstacle with a not exactly parallel trajectory, the orthogonal assumption is used to correct those slightly deviated segments. Oblique segments are allowed but not changed. This means a significant improvement to the mapping process because, as Yamauchi points out [10], it is still common to find in the literature systems that can only handle parallel or perpendicular walls.

3 MAP GENERATION

The space being explored by the robots is discretised by means of a grid. Cells in the grid represent a small area of the real environment and contain two values: the degree of possibility and the degree of necessity of the presence of obstacles. Initially, that is, before any exploration has taken place, each cell, represented by its co-ordinates (x_i, y_j), has a possibility value $\Pi_{ij}(\text{wall})=1$ (i.e., it is completely possible that there is a wall or obstacle in the cell) and a necessity value $N_{ij}(\text{wall})=0$ (i.e., there is no certainty at all that there is a wall or obstacle in the cell). These initial values correspond to a situation of total ignorance according to the theory of possibility [2]. As robots communicate the information gathered during their exploration to the host, the possibility and necessity values are modified in a way that depends on the detection, or not, of obstacles by the returning robot. The information gathered by each robot is the trajectory of the robot together with the position of the walls and singular points (that is wall ends and corners) that have been detected along it. Due to the odometry error, the position of the detected walls has an associated error. This error follows a Normal distribution [5] and has been approximated by a rectangle whose size increases with the travelled distance.

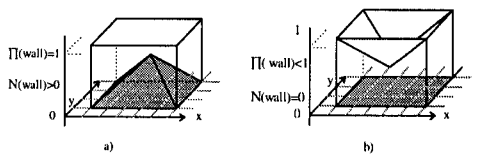


Figure 1: Π and N values assigned to cells corresponding to: a) wall detection, and b) free space.

When an error rectangle is associated to a position that belongs to a detected wall, the occupancy certainty degree (i.e., the certainty about the presence of an obstacle in that position) is expressed by means of necessity values in every cell that results covered by the error rectangle around that position. The necessity values decrease linearly (from the centre of the error rectangle) with the

magnitude of the error and remains positive (N_{ij}(wall) > 0) for all the cells inside the error rectangle but gets the value 0 at the cells outside the limits of the rectangle. Figure 1 a) shows this case. Notice that according to the axioms of possibility theory, the possibility value is constantly equal to 1 in all the cells covered by the error rectangle.

On the other hand, robot trajectories supply information about free space so that the corresponding cells (covered by the error rectangle) have values of $\Pi_{ij}(\neg wall) = 1$ and $N_{ij}(\neg wall) > 0$. According to the axioms of possibility theory, these values are equivalent to $N_{ij}(wall) = 0$ and $\Pi_{ij}(wall) < 1$. Figure 1 b) shows this case, and the next subsection describes how these possibilities and necessities are computed.

3.1 VALUE ASSIGNMENT AND COMBINATION

The height of the pyramids in Figure 1 are determined by the size of the error rectangle. The underlying idea is to establish a linear error-height relation such that, a null error implies the maximum allowed value (1) of height, and an error equal to the maximum allowed error (K) implies a zero height since the information is no longer reliable.

Let us represent the error rectangle by the tuple (x_c, y_c, $error_x$, $error_y$), where x_c, and y_c are the coordinates of the central cell of the rectangle and $error_x$ and $error_y$ are, respectively, half the length of the base and half the length of the height of the error rectangle measured in number of grid cells. Following this, the height of the pyramid is given by:

$$height = 1 - \frac{Max(error_x, error_y)}{K},$$

and the necessity value of having a wall at (x_i, y_j) is:

$$N_{ij}(wall) = height * min(1 - \frac{|x_i - x_c|}{error_x}, 1 - \frac{|y_j - y_c|}{error_y})$$

and the Π_{ij}(wall) is obtained through 1-N_{ij}(¬wall) by using the same equations.

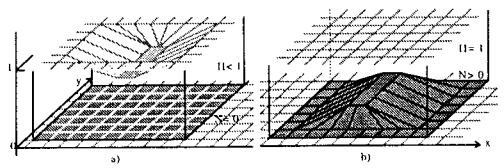


Figure 2: Segment representation corresponding to a) trajectories (possibility pyramids), and b) walls (necessity pyramids).

In considering now consecutive points along the trajectory of the robot or along a wall segment, since their corresponding pyramids overlap, some value combination is required. In the case of wall segments the values are necessities (increasing from 0) and are combined by using the *max* operation (figure 2 b)). In the case of trajectories these values are possibilities (decreasing from 1) and are combined by means of the *min* operator (Figure 2 a))

When the same portion of a wall has been followed more than once, we consider them as independent sources of evidence. Thus, the necessities are combined by means of the probabilistic sum, that is $S(x,y) = x+y-x \cdot y$, in order to reinforce the certainty about the location of the wall. Figure 3 shows a top view of this situation.



a) b)

Figure 3: a) First partial detection of a wall, b) the same wall is detected during a different trajectory segment.

Although the definition of our possibility and necessity pyramids are based on a statistical analysis of the displacement error, the use of Possibility Theory allows us to define ignorance as well as to provide a simple value combination (and reinforcement).

4 MAP EXTENSION

Following a wall along a random distance implies that very often the robots leave the wall before reaching its end. This is done in order to increase the number of discovered features and to avoid that robots keep following the contour of a feature once it is detected. If it is the case that the end is reached, then the corresponding extreme of the segment is labelled as "singular point"). Since we can ensure that wall segments without singular points correspond to longer walls —or obstacle edges— they can be extended. However, we cannot know their real length and therefore we make the following decisions regarding their limits: robot trajectory segments limit the extension of wall segments, we stop extending a wall segment when it meets either a detected wall segment or another segment extension (in the former case they belong to the same wall, while the later represents perpendicular walls).

Extension is done locally by propagating small constant certainty values (they are just assumptions) at the ends of orthogonal wall segments. Since, oblique segments represent shorter features as doors, or other less predictable objects, they are not extended.

Wall extension has two main advantages: first, it increases the coverage of the real environment and second, the planning [8] over the resulting maps give more conservative and safer paths than the ones obtained considering just detected features. These paths are computed from a visibility graph of the free space and guide robots towards less explored areas, reducing the use of reactivity (to avoid non previously detected obstacles) when following the paths.

5 RESULTS

Five robots have been sent to explore an unknown environment (see Fig 4 a)). Since the host builds its map (Fig 4 b)) in an incremental way, the data in Table 1 shows the environment coverage each time one more robot delivers its exploration information to the host. Columns list the wall coverage percentage (respect a total length 30m.), first column stands for detected wall segments, second for correctly extended wall segments and the third represents the incorrect extensions.

Table 1: percentage of the environment coverage

# maps	*coverage*	*extension*	*incorrect ext.*
1	16.3%	32%	1.97%
2	28.2%	48.99%	1.97%
3	44.6%	58.7%	0.4%
4	58.8%	78.1%	-
5	65.2%	82.1%	-

The extension process is purely based on the application of heuristics and, therefore, it is hard to evaluate. It is also highly dependent on the structure of the environment as well as on the detections from random explorations.

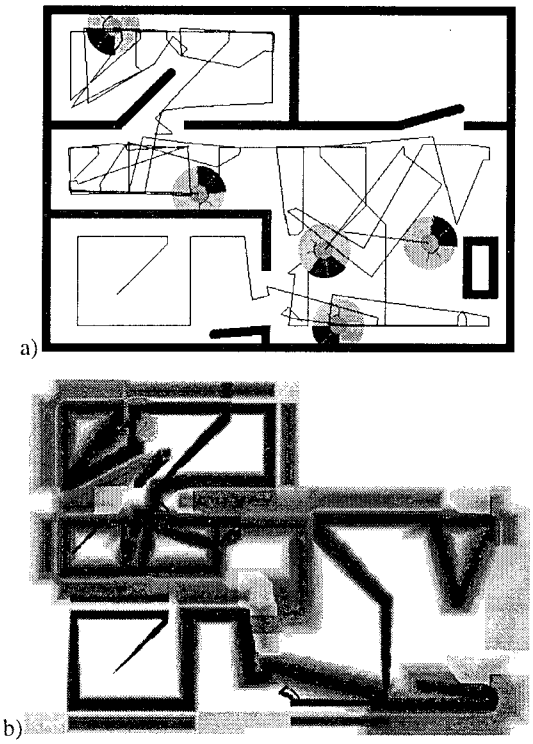


Figure 4: a) Robots explore an unknown 5.3×3.8 m. environment b) Extended map: wall detections in red, trajectories in blue, singular points in green and wall extensions in light red (darker colours mean more certainty).

In order to further illustrate the improvement, we have applied the same heuristics to the environment at Fig 5 a). In this case, the detections are shown in Fig 5 b) and cover 45.3% of the total walls. Fig 5 c) depicts the extensions, from which 70.1% are correct and 6.4% are incorrect. This second example has been also used to compare the planning in extended and non-extended maps. Fig 5 b) shows planned paths in the non-extended map and c) the paths in the extended map. Notice that paths planned on extended maps are named after those in the non-extended map, but a letter 'e' differentiates them. Obviously, they have the same initial and goal position.

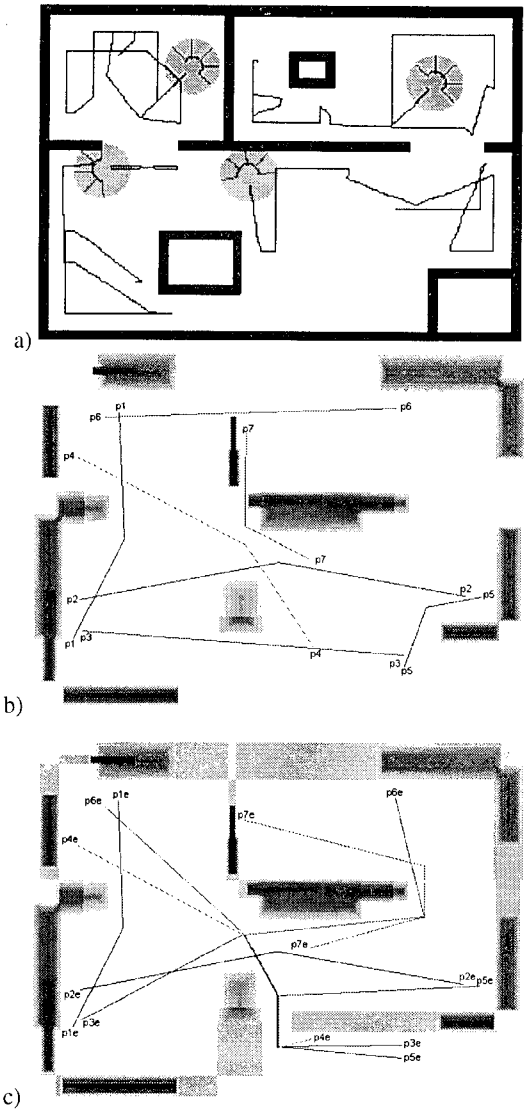


Figure 5: a) Robot exploration . Paths planned on the resulting map considering: b) wall detections and c) extensions.

The three images in Figure 6 depict the trajectories actually followed by the robots –some of them present common trajectories. Obviously, in areas with the same information both perform equally (see path p1 and p2). When one map is more accurate than the other, incorrectly extended information yields to paths longer than necessary, while non-detected walls imply a high use of reactivity. Sometimes, an incorrect extension closes narrow gaps that require the use of reactive capabilities to get through. In that way, an initially shorter path in the non-extended map can result in a trajectory whose length is in fact comparable to the length of a trajectory from the extended map (this is the case of p3 and p3e). Of course, this is not always the case: p4e avoids an actually non-existing piece of wall causing an extra displacement of 0.37m. (p4 is a 6.45m long trajectory). The worse case of avoiding a non-existent wall is p5e, which turns around the corner where a previous robot went. Although this tendency of going over previous robot trajectories can be a disadvantage, already explored free space is also the source of safer paths. Finally, when extension is correct,

resulting paths are more informed and therefore reactivity is less often needed (see p6 & p6e).

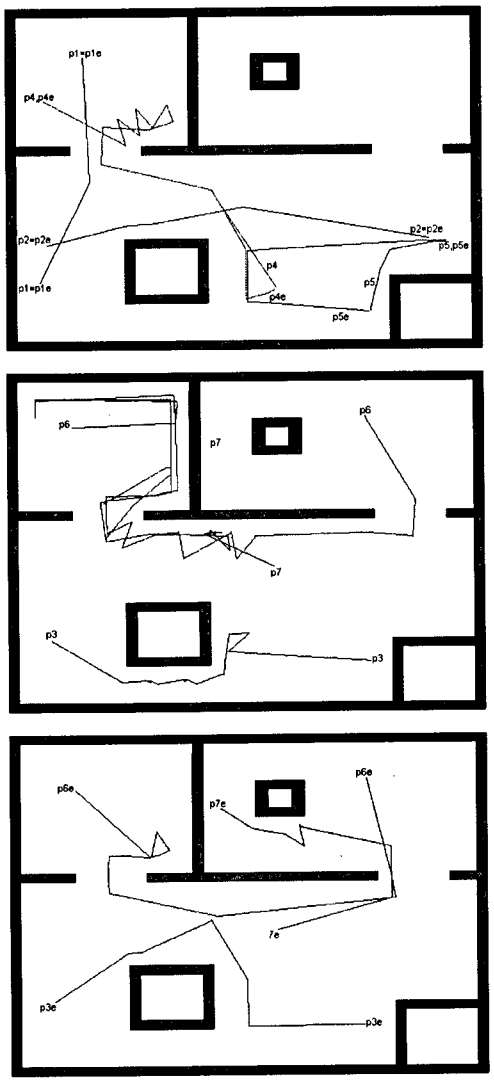


Figure 6: Performance of the robots following the paths at Figure 5.

Reactivity is less efficient and can yield the problem that appears in p7, in which the reactivity makes the robot to take so many wrong decisions that the robot stops before reaching the target (the robot position error grows so much due to the many wrong displacement movements that the area it covers already includes the target co-ordinates even when the robot is still quite far from the target).

5 RELATED WORK

Probabilistic techniques are commonly used for map representation, for example [1] defines landmarks in natural environments assuming a Gaussian certainty distribution of their positions and [4,6,7] estimate the probability of cell occupancy in certainty grids. Probabilistic techniques need a huge amount of data and assume known distributions, Fuzzy Set theory is a good alternative when these conditions are not met. For example [3] uses fuzzy numbers to model the uncertainty of the parameters of geometric primitives and of co-ordinate transformations used to describe natural environments and [9] uses fuzzy sets to assign to each point in the grid map a degree of being empty and a degree of being occupied. This last work is the closest to ours, but our use of infrared sensors in short distances gives better results than their imprecise information coming from ultrasonic sensors.

References

[1] Betgé-Brezetz, Hébert, Chatila, Devy, Uncertain Map Making in Natural Environments, Proc *IEEE: Int. Conf. on Robotics and Automation*, 1048-1053, 1996.

[2] Dubois and Prade, *Possibility Theory*, Plenum Press, 1988.

[3] Kim, Ko and Chung, Uncertain robot environment modelling using fuzzy numbers, *Fuzzy Sets and Systems*. Vol 61, pp 53-62, 94.

[4] Lim, and Cho, Physically Based Sensor Modelling for Sonar Map in Specular Environment, Proc *IEEE Conf on Robotics and Automation*, pp 1714-1719, 1992

[5] López-Sánchez, López de Màntaras and Sierra, Incremental map generation by low cost robots based on possibility/necessity grids, P. *Uncertainty in Artificial Intelligence*, pp. 351-357, 1997.

[6] Movarec and Elfes, High Resolution Maps from Wide Angle Sonar, P. *Int. Conf. on Robotics and Automation*, pp 116-121, 1985.

[7] Pagac, Nebot and Durrant-White, An evidential approach to probabilistic map-building, Proc *IEEE Int Conf on Robotics and Automation*, pp 745-750, 1996.

[8] Pearl, *Heuristics: Intelligent Search Strategies for Computer Problem Solving*, Addison-Wesley. 1984.

[9] Poloni, Ulivi and Vendittelli, Fuzzy logic and autonomous vehicles: Experiments in ultrasonic vision, *Fuzzy Sets and Systems*. Vol 69, pp15-27, 1995.

[10]Yamauchi, B. Frontier-based exploration using multiple robots, Proc *Autonomous Agents*, 47-53,1998.

THEORY AND APPLICATION OF MULTIPLE-VALUED LOGICS FOR KNOWLEDGE-BASED SYSTEMS

José Antonio Reyes **Francesc Esteva** **Josep Puyol-Gruart**
Artificial Intelligence Research Institute, IIIA
Spanish Council for Scientific Research, CSIC
Campus UAB. 08193 Bellaterra, Catalonia, Spain
E-mail: {reyes, esteva, puyol}@iiia.csic.es

Abstract

Multiple-valued logics are useful for dealing with uncertainty and imprecision in knowledge-based systems. In this paper, we present the foundations of **QMORPH**: a tool that assists users in the declaration of such logics and in the declaration of the communication mechanism between two of these different logics by preserving inference.

Keywords: uncertain reasoning, multiple-valued logics, knowledge-based systems.

1 DEFINING MV-LOGICS

Multiple-valued logics (MV-Logics) have been proved to be a useful way to manage uncertainty and imprecision [3; 5; 6; 10].

We consider a restricted family of finite MV-logics expressive enough to model the uncertain reasoning used in many rule-based systems [1]. Each logic is determined by a particular algebra of truth-values from this family.

An **Algebra of truth values** is a finite algebra $A_T^n = \{A_n, \mathbf{0}, \mathbf{1}, N_n, T, I_T\}$ such that:

1. The ordered set of **truth-values** A_n is a chain of n linguistic terms:

$$\mathbf{0} = a_0 < a_1 < \ldots < a_{n-1} = \mathbf{1}$$

where $\mathbf{0}$ and $\mathbf{1}$ are the Boolean *False* and *True* respectively.

2. The **negation** operator N_n is the unary operation defined as:

$$N_n(a_i) = a_{n-1-i}$$

the only one that fulfills the following properties:

N1: If $a < b$ then $N_n(a) > N_n(b)$, $\forall a, b \in A_n$

N2: $N_n^2 = Id$

3. The **conjunction** operation T is any binary operation such that the following properties hold $\forall a, b, c \in A_n$:

T1: $T(a,b) = T(b,a)$
T2: $T(a,T(b,c)) = T(T(a,b),c)$
T3: $T(\mathbf{0},a) = \mathbf{0}$
T4: $T(\mathbf{1},a) = a$
T5: If $a \leq b$ then $T(a,c) \leq T(b,c)$ for all c

4. The **implication** operation I_T is defined by residuation with respect to T, i.e.

$$I_T(a,b) = Max\{c \in A_n \mid T(a,c) \leq b\}$$

Therefore, to establish an algebra of truth-values it is only necessary to determine the set of truth-values A_n more adequate for the concerning problem, and define a conjunction operator T.

The **sentences** of the language are pairs (p, V), where: p is a wff, obtained in the usual way from a denumerable set of propositional symbols and the connectives *'not'* ($\neg$), *'and'* ($\wedge$) and *'detachment'* ($\rightarrow$); and V is an interval of truth-values from A_n.

Likewise, semantic is evaluated in an algebra of truth-values. The semantic interpretation is given by the following signification:

- **Models** are defined by valuations; i.e. mappings ρ from the propositional symbols to A_n terms provided that:

$$\rho(\neg p) = N_n(\rho(p))$$
$$\rho(p_1 \wedge p_2) = T(\rho(p_1), \rho(p_2))$$
$$\rho(p \rightarrow q) = I_T(\rho(p), \rho(q))$$

- The **Satisfaction Relation** between models and sentences is defined by:

$M_\rho \models (p, V)$ if, and only if, $\rho(p) \in V$,

where M_ρ stands for the model defined by a valuation ρ.

The inference rule of these logics is *Modus Ponens*, which gives from $(p \rightarrow q, V)$ and (p, V') a sentence (q, V''). In fact, modus ponens evaluates V'' from V and V' (See [9] for a theoretical study of modus ponens, and [6] for its application to expert systems).

1.1 INTERVALS OF TRUTH-VALUES

We extend the previous algebra to an algebra of intervals, [4]. We have three motivations to do this. The first one is imprecision. The second reason is related with the modus ponens operator, which requires the use of intervals to

chain rules. And finally, intervals are needed to make possible mappings between different logics.

Given an algebra of truth-values $A = < A_n , N_n , T >$, we will consider the set of intervals of A_n as:

$$Int(A_n) = \{[a,b] \mid a,b \in A_n\}$$

being $[a,b] = \{x \in A_n \mid a \leq x \leq b\}$.

The extensions to intervals of the above operators are: $N_n^*([a,b]) = [N_n(b), N_n(a)]$, $T^*([a,b], [c,d]) = [T(a,c), T(b,d)]$.

1.2 CONJUNCTION GENERATION

Given a set of linguistic terms, we need to know all the possible conjunction operators in order to choose one of them.

The mentioned algebra properties (**T1-T5**) are requirements that a conjunction operator must fit. We have modelled this task as a classic CSP problem where these properties act as constraints over the set of possible solutions. We represent the conjunction operator as a matrix, where each element is a variable $V_{i,j}$, $0 \leq i, j \leq n-1$. Satisfaction of each property or constraint, causes the following guidelines which influence the conjunction search generation problem:

- Commutativity allows us to consider only the set of variables $V = \{V_{i,j}, i \geq j\}$.
- The absorbent and neutral elements fix the values for variables $V_{0,j}$ and $V_{j,n-1}$.
- Monotonicity implies non-decreasing rows and columns.
- Associativity test is expensive in time and memory, but considering that $T(a_i,a_j) \leq min(a_i,a_j)$, each sub-matrix where $i = j$ is closed with respect to the conjunction operator. In this way, the matrix will be associative if all its sub-matrices are so. This property lets us check associativity in an incremental way every time a value is assigned to a variable $V_{i,i}$.

Applying these constraints, the number of operators generated is exponential with respect to n. We can consider two options to reduce this number: introducing some new desirable constraints, as well as fixing several values of the matrix to reduce the search space to be explored.

Two types of additional, and optional, constraints are discussed:

T6 Strictness: $T(a_i,a_j) \neq 0$, for all $i, j \neq 0$

Given two truth-values different than 0 *(false)*, it does not seem reasonable that their conjunction may be 0.

T7 α-Smoothness: given $\alpha \in \mathbb{N}$, T is said to satisfy α-smoothness property if $T(a_i,a_j) = a_k$ and $T(a_{i-1},a_j) = a_p$, then $k - p \leq \alpha$

We also want that the result of the conjunction of near values has to be not very distant.

Satisfaction of these new properties (**T6, T7**) has the following consequences in the conjunction search generation problem:

- Strictness: the generation of strict matrices of dimension n is equivalent to generate non-strict matrices of dimension $n-1$.
- α-Smoothness: it reduces the set of possible values for a variable taking into account column and row adjacency.

2 COMBINING LOGICS

We can declare different logics by varying the set of truth-values (linguistic terms) and the conjunction operator. That depends of how the expert will deal with uncertainty in each problem.

When two different logics need to exchange information, it is necessary some mechanisms of translation to make the communication between these logics compatible. We want to preserve the inference properties of each logic but, as usual, it is not possible to transmit information without loss of precision. Hence we map values of an algebra into intervals of the other.

Let's take a look to which requirements we need in order to map the language of a logic into another one when they require to exchange information [2].

2.1 MAPPINGS BETWEEN MV-LOGICS

Let $(L, \vdash)$ and $(L', \vdash')$ be two logics, L and L' standing for the languages, $\vdash$ and $\vdash'$ for the entailment relations defined on L and L' respectively. To establish a correspondence between both logics, a *mapping* $H: L \longrightarrow L'$ is needed. Now, we will analyze some natural requirements for the mapping H with respect to the entailment systems $\vdash$ and $\vdash'$. We propose that at least one of the following three requirements should be fulfilled by the mapping H in order to ensure a consistent communication. Henceforth Γ and e will denote a set of sentences and a sentence of L respectively. A map H is said to be a *forward conservative* map when,

RQ.1. If $\Gamma \vdash e$, then $H(\Gamma) \vdash' H(e)$

For every sentence e, deducible from a set of sentences Γ, its corresponding sentence, $H(e)$, will also be deducible from the corresponding sentences of $H(\Gamma)$.

A map fulfilling this second requirement is said to be a *backward conservative* map:

RQ.2. If $H(\Gamma) \vdash' H(e)$, then $\Gamma \vdash e$

This is the inverse requirement of **RQ.1**. Hence, if a fact is not deducible from Γ, then its corresponding fact from $H(\Gamma)$ won't be deducible either. Nevertheless $\Gamma \vdash e$ does not imply $H(\Gamma) \vdash' H(e)$.

Conditions **RQ.1** and **RQ.2**, which are very strong, can sometimes be weakened in the uncertain reasoning framework. Formally, this can be expressed by the third and last requirement:

RQ.3. If $H(\Gamma)\vdash' e'$, then there exists e such that $\Gamma\vdash e$ and $H(e)\vdash' e'$

This requirement assures that every sentence deducible from $H(\Gamma)$ must be in agreement with what can be deduced from Γ. This requirement is slightly different from **RQ.2**, in the sense that it is not necessary e' being an exact translation of a deducible sentence e from Γ, but only something deducible from such a translation. In the framework of logics for uncertainty management, e' is interpreted as a *weaker* form of e, i.e. a sentence expressing more uncertainty than e. We will call it a *weak conservative* map.

Now we will consider the problem of finding inference preserving correspondences between two logics $A = <A_n , N_n , T >$ and $B = <B_m , N_m , T'>$. We are interested in mapping the entailment system $(L_A,\vdash_A)$ into the entailment system $(L_B,\vdash_B)$, by means of renaming functions between the corresponding linguistic term sets. This means that we will only consider those mappings translating sentences from L_A to L_B that just involve translations of truth-values; i.e. any mapping $H: L_A \longrightarrow L_B$ will be defined as $H(e,V) = (e,h(V))$, where h translates subsets of values of A_n into subsets of values of B_m, and e (a wff) remains invariant.

2.2 ALGEBRA MORPHISMS

To establish conservative communications, it is necessary to consider what kind of relation between the uncertainty logics is required. In [2], we can find the necessary and/or sufficient conditions for a mapping $h: A_n \longrightarrow Int(B_m)$ to satisfy these requirements. From this analysis, we deduce the following relationships between the conditions needed to satisfy every requirement (see Figure 1):

$$\textbf{RQ.1} \Leftrightarrow \left\{\begin{array}{c} h(T(V1,V2)) \supseteq T'(h(V1),h(V2)) \\ h(N(V)) = N_m(h(V)) \\ h(V1 \cap V2) = h(V1) \cap h(V2) \end{array}\right\}$$

$$\textbf{RQ.2} \Leftarrow \left\{\begin{array}{c} h(T(V1,V2)) \subseteq T'(h(V1),h(V2)) \\ h(N(V)) = N_m(h(V)) \\ h(V1) \supseteq h(V2) \Rightarrow V1 \supseteq V2 \end{array}\right\}$$

$$\Downarrow$$

$$\textbf{RQ.3} \Leftrightarrow \left\{\begin{array}{c} h(T(V1,V2)) \subseteq T'(h(V1),h(V2)) \\ h(N(V)) = N_m(h(V)) \end{array}\right\}$$

Figure 1: Requirements conditions.

We define mappings from elements of A_n into intervals of B_m, but sometimes it is possible to find mappings that translate an element of A_n into an element of B_m. In this case, requirement **RQ.3** is satisfied and the mapping is a *morphism* between the corresponding algebras. As a particular case, if a map fulfill **RQ.1** and **RQ.2**, we have not only the morphism conditions, but also a one-to-one application, that is, an injective function called *monomorphism*. But we can not always find these kinds of mappings, so in the case of a map involving intervals of truth-values, we call it a *quasi-morphism*.

We are mainly interested in *monomorphisms* because they embed A_n into B_m and because they are order preserving mappings (it is equivalent to a communication without any loose of information). The second preference is *morphisms*, which accomplish the algebra operations (it is a transmission of information fulfilling the required properties). Finally, because of the strong conditions *morphisms* and *monomorphisms* must satisfy, it is not always possible to find these kind of renaming them, so *quasi-morphisms* can be useful thanks to the additional freedom of map truth-values of an algebra into intervals of the other (we allow certain loose of information).

2.3 RENAMING GENERATION

For a given set of truth-values A_n, there exists only one negation operator N_n. Then, we can make the following partition:

- the set of negative elements $\boldsymbol{N_n} = \{x \mid x < N_n(x)\}$
- the set of fixed elements $\boldsymbol{F_n} = \{x \mid x = N_n(x)\}$
- the set of positive elements $\boldsymbol{P_n} = \{x \mid x > N_n(x)\}$

Besides, we can do the same with a set B_m, obtaining $\boldsymbol{N_m^*}$, $\boldsymbol{F_m^*}$ and $\boldsymbol{P_m^*}$ for the case of working with intervals. Then, the renaming algorithm consists in generating all the maps h_I: $\boldsymbol{N_n} \cup \boldsymbol{F_n} \longrightarrow \boldsymbol{N_m^*} \cup \boldsymbol{F_m^*}$ such that:

a) $h_I(0) = 0$, (i.e. $h_I(a_0) = b_0$).

b) $h_I(\boldsymbol{F_n}) \in \boldsymbol{F_m^*}$.

c) $x \leq y$ implies $h_I(x) \not> h_I(y)$, where $x, y \in A_n$ and $h_I(x), h_I(y) \in Int(B_m)$.

and extending each of these mapping h_I with respect to the negation operation defining the morphism h as:

$$h(x) = \begin{cases} h_I(x), & \text{if } x \in \boldsymbol{N_n} \cup \boldsymbol{F_n}. \\ N_m^*(h_I(N_n(x))), & \text{if } x \in \boldsymbol{P_n}. \end{cases}$$

Next, we check which ones are compatible with the conjunction operators T and T', and finally, we check which ones are monomorphisms, morphisms or quasi-morphism. The resulting maps are presented to the user in this order.

Due to the strong condition monomorphisms and morphisms must fit, it is not always possible to find them. However, it is very possible that the renaming generation produces a large list of quasi-morphisms, so the possibility of giving an ordered list of quasi-

morphisms to aid users in their selection may be considered. Note that in the case of morphisms, all the renaming mappings have the same evaluation, hence the selection is left to the user's criteria.

For the purpose of producing an ordered list of quasi-morphisms, we consider a weigh among the cardinal c of the set of terms which have an atomic image (that is, an interval with the form $[a_i, a_i]$) and the length L of the remainder of intervals (number of truth-values included in this interval).

Given two chains A_n and B_m, we will generate mappings between A_n and the set of intervals $Int(B_m)$. In order to obtain an evaluation for every map, we can use the following empirical function:

$$\eta = \frac{\sum_{i=1}^{n} L_i}{\sum_{i=1}^{n} c_i}, \; 1 \le L_i \le m, \; 0 \le c_i \le 1 \qquad (1)$$

where L_i is the length of the interval i and c_i is the number of points which have an atomic image.

User is also able to establish the behavior that can take the generated maps defining a partial map between both algebras. This definition must satisfy the set of requirements suggested previously in order to be a negation morphism. When we have defined our partial map (or total, or none), a list of renaming functions fulfilling these defined characteristics will be generated.

3 THE QMORPH TOOL

QMORPH [8] is a tool that allows users to define the more adequated logic for a concrete problem, as well as to decide a consistent communication between two different MV-logics. This tool makes automatic these problems and assists users along the process. Two different interfaces have been developed: a graphical interface for the UNIX operative system using *Tcl/Tk* packages, running under the X-Windows environment and developed on Sun machines; and a generic text mode interface performed in Common Lisp.

In the present, this tool is attached to **Milord II** [7] - a specific expert system building environment - as an aid tool to support experts when developing modular applications, although it can be used in a more extensive framework.

4 CONCLUSIONS

In this paper we have defined which are the theoretical bases that allow dealing with uncertainty by means of multiple-valued logics. We have settled how to declare a suitable MV-logic from a parametric family of algebra of truth-values, and also, which are the necessary and/or sufficient requirements when we need to establish a communication between two of these logics. This problem arises in large knowledge-based systems in which different tasks need to cooperate using uncertain reasoning, as well as in distributed systems.

QMORPH has been designed with the aim of automating these problems, as well as assisting users and offering different alternative possibilities. Beyond its particular use within **Milord II**, it can be deployed by any other system using MV-logics.

This work focuses on the unidirectional interaction between two logics, but the more general problem of communicating various uncertain reasoning systems is far more complex. This work along with further research will make possible this goal.

Acknowledgments

Partially supported by project MODELOGOS funded by CICYT (TIC 97-0579-C02-01) and through SMASH by CICYT (TIC 96-1038-C04-01).

References

[1] Agustí, J.; Esteva, F.; Garcia, P.; Godo, L.; Sierra, C.; Combining multiple-valued logics in modular expert systems, *Proceedings of 7th Conference on Uncertainty in AI*, Bruce d-Ambrosio et al. (eds), (1991), pp. 17-25.

[2] Agustí, J.; Esteva, F.; Garcia, P.; Godo, L.; López de Mántaras, R.; Local multi-valued logics in modular expert systems, *Journal of Experimental & Theoretical AI (JETAI)*, vol. 6 n. 3 (1993), pp. 303-321.

[3] Bonissone, P.; Gans, S.; Decker, K.; Rum: A layered architecture for reasoning with uncertainty, in *IJCAI'87* (1987), pp. 891-898.

[4] Esteva, F.; Garcia-Calves, P.; Godo, L.; Enriched interval bilattices: An approach to deal with uncertainty and imprecision, *International Journal on Uncertainty, Fuzzyness and Knowledge-Based Systems*, 1-2 (1994).

[5] Godo, L.; López de Mántaras, R; Sierra, C.; Verdaguer, A.; Milord : The architecture and management of linguistically expressed uncertainty, *International Journal of Intelligent Systems*, Vol. 4 n. 4 (1989), pp. 471-501.

[6] López de Mántaras, R.; *Approximate Reasoning Models*, Ellis Horwood Series in Artificial Intelligence, 1990.

[7] Puyol, J.; *MILORD II: A language for knowledge-Based Systems*, Vol. 1 of Monografies del IIIA, IIIA-CSIC, 1996.

[8] Reyes, J. A.; **QMORPH**: A tool to define and combine local logics in Milord II, *Mst. Thesis, Universitat Autònoma de Barcelona*, (1997).

[9] Trillas, E.; Valverde, L.; On mode and implication in approximate reasoning, in Gupa et al. (eds.): *Approximate Reasoning in expert systems*, (1985), pp. 157-166.

[10] Turner, R.; *Logics for Artificial Intelligence*, Ellis Horwood Series in Artificial Intelligence, 1984.

Axiomatizing t-Norm Based Logic

Siegfried Gottwald
Leipzig University, D-04109 Leipzig, Germany
gottwald@rz.uni-leipzig.de

Abstract

In fuzzy logic in wider sense, t-norms got a prominent rôle in recent times.

In many-valued logic, the LUKASIEWICZ systems, the GÖDEL sytems, and also the product logic are t-norm based. The present paper discusses the more general problem of the adequate axiomatizability for such t-norm based logical systems in general, surveying results of the last years.

Keywords: fuzzy logic, many-valued logic, t-norms, t-norm based connectives.

1 Semantic approach

There is an important difference to the standard approach toward semantically based systems of many-valued logic: in the area of t-norm based systems there is no single, "standard" semantical matrix for the general approach.

The most appropriate way out thus shall be: to find some suitable class(es) of algebraic structures which can be used to characterize these logical system(s).

An important restriction has to be met: to left continuous (or even to continuous) t-norms, because one (usually) likes to have an R-implication connective together with a basic t-norm.

As it seems, the notions of left continuity and of continuity need the reference to the notion of limit. And this is not (genuinely) an algebraic notion.

Hence it would be nice to find – algebraically characterizable – classes of algebraic structures which "approximate" well the class of algebraic structures determined by the continuous, or by the left continuous t-norms.

The following is structurally important for t-norms:

- $\langle [0,1], t, 1\rangle$ is a commutative semigroup with a neutral element, i.e. a commutative monoid,
- $\leq$ is a (lattice) ordering in $[0,1]$ which has a universal lower bound and a universal upper bound,
- both structures "fit together": t is non-decreasing w.r.t. this lattice ordering.

Thus one has to consider (complete) abelian lattice-ordered monoids as the truth degree structures for the t-norm based systems.

In general, abelian lattice-ordered monoids may have different elements as the universal upper bound of the lattice and the neutral element of the monoid. This is not so for the t-norm based systems, they have an *integral* abelian lattice-ordered monoid as truth degree structure.

One also likes to have the t-norm combined with a pseudo-complementation, its R-implication operator: i.e. the abelian lattice-ordered monoid formed by the truth degree structure has also to be a *residuated* one.

At all, hence, we consider *residuated lattices*, i.e. algebraic structures $\langle L, \cap, \cup, *, \rightarrowtail, 0, 1\rangle$ such that L is a lattice under $\cap, \cup$ with zero 0 and unit 1, and an abelian lattice-ordered monoid under $*$ with neutral element 1, and such that the operations $*$ and $\rightarrowtail$ form an adjoint pair, i.e. satisfy

$$z \leqslant (x \rightarrowtail y) \quad \Leftrightarrow \quad x * z \leqslant y.$$

It is nice to recognize that the adjointness condition is in the present setting the suitable algebraic equivalent of the analytical notion of left continuity.

Proposition 1 *For any t-norm $\boldsymbol{t}$ one has that $\boldsymbol{t}$ and its R-implication form an adjoint pair iff $\boldsymbol{t}$ is left continuous (in both arguments).*

And also the continuity of the basic t-norm has an algebraic equivalent: the property of divisibility for the commutative lattice-ordered monoids.

Definition 1 $\langle L, *, 1, \leqslant \rangle$ *is* divisible *iff for all $a, b \in L$ with $a \leqslant b$ there exists some $c \in L$ with $a = b * c$.*

For residuated lattices one has another characterization of divisibility:
$\langle L, \cap, \cup, *, \rightarrowtail, 0, 1 \rangle$ is divisible iff one has $a \cap b = a * (a \rightarrowtail b)$ for all $a, b \in L$.

Proposition 2 *A t-norm based residuated lattice $\langle [0,1], \min, \max, \boldsymbol{t}, \mathsf{seq}_{\boldsymbol{t}}, 0, 1 \rangle$ is divisible iff $\boldsymbol{t}$ is continuous.*

A further restriction is suitable because each t-norm based residuated lattice is linearly ordered, and thus makes the wff $(\varphi \rightarrow_{\boldsymbol{t}} \psi) \vee (\psi \rightarrow_{\boldsymbol{t}} \varphi)$ valid.

So finally we call *BL-algebras* divisible residuated lattices which also satisfy the *pre-linearity* condition:
$(x \rightarrowtail y) \cup (y \rightarrowtail x) = 1$.

2 Syntactic approach

Now we consider the problems of syntactic characterizations of the classes of all wffs which are valid in all residuated lattices, or in all BL-algebras. For more details the reader may e.g. consult [2].

2.1 Monoidal logic

The axiom system $\mathsf{Ax}_{\mathsf{ML}}$ of HÖHLE [6, 7] for the class of wffs which are valid in all residuated lattices has the axiom schemata:

$$(\varphi \rightarrow \psi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\varphi \rightarrow \chi)),$$
$$\varphi \rightarrow \varphi \vee \psi,$$
$$\psi \rightarrow \varphi \vee \psi,$$
$$(\varphi \rightarrow \chi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\varphi \vee \psi \rightarrow \chi)),$$
$$\varphi \wedge \psi \rightarrow \varphi,$$
$$\varphi \,\&\, \psi \rightarrow \varphi,$$
$$\varphi \wedge \psi \rightarrow \psi,$$
$$\varphi \,\&\, \psi \rightarrow \psi \,\&\, \varphi,$$
$$(\varphi \,\&\, \psi) \,\&\, \chi \rightarrow \varphi \,\&\, (\chi \,\&\, \chi),$$
$$(\varphi \rightarrow \psi) \rightarrow ((\varphi \rightarrow \chi) \rightarrow (\varphi \rightarrow \psi \wedge \chi)),$$
$$(\varphi \rightarrow (\chi \rightarrow \chi)) \rightarrow (\varphi \,\&\, \chi \rightarrow \chi),$$
$$(\varphi \,\&\, \psi \rightarrow \chi) \rightarrow (\varphi \rightarrow (\psi \rightarrow \chi)),$$
$$\varphi \,\&\, \neg\varphi \rightarrow \psi,$$
$$(\varphi \rightarrow \varphi \,\&\, \neg\varphi) \rightarrow \neg\varphi,$$

and has as its (only) inference rule the rule of detachment (w.r.t. the implication connective $\rightarrow$).

The logical calculus which is constituted by this axiom system and its inference rule, and which has the standard notion of derivation, shall be denoted by $\mathbb{K}_{\mathsf{ML}}$.

Proposition 3 *The calculus $\mathbb{K}_{\mathsf{ML}}$ is sound, i.e. proves only such formulas which are valid in all residuated lattices.*

Corollary 4 *The* LINDENBAUM *algebra of the calculus $\mathbb{K}_{\mathsf{ML}}$ is a residuated lattice.*

Theorem 5 (Completeness) *For a wff H of $\mathcal{L}_{\boldsymbol{t}}$ there are equivalent:*
(i) H is derivable within the logical calculus $\mathbb{K}_{\mathsf{ML}}$;
(ii) H is valid in all residuated lattices.

Proof: The part $(i) \Rightarrow (ii)$ here is just the soundness of the calculus $\mathbb{K}_{\mathsf{ML}}$. The part $(ii) \Rightarrow (i)$ on the other hand results from the fact that the LINDENBAUM algebra of $\mathbb{K}_{\mathsf{ML}}$ is a residuated lattice and has the class of all $\mathbb{K}_{\mathsf{ML}}$-derivable wffs as its unit element, i.e. as the universal upper bound of its lattice part and at the same time as the neutral element of its monoidal part. Therefore each non-$\mathbb{K}_{\mathsf{ML}}$-derivable wff is not valid in at least one residuated lattice. □

This monoidal logic can be specialized toward quite different systems of (non-classical) logic by adding suitable further axioms.
Remark: The condition of pre-linearity is satisfied in all MV-algebras.

Theorem 6 *From the monoidal logic one gets adequate axiomatizations of:*
(a) the ŁUKASIEWICZ *system* L_∞ *if one adds the axiom schemata*

$$\varphi \wedge \psi \to \varphi \,\&\, (\varphi \to \psi)\,,$$
$$\neg\neg\varphi \to \varphi\,;$$

(b) the GÖDEL *system* G_∞ *if one adds the axiom schemata*

$$\varphi \to \varphi \,\&\, \varphi\,,$$
$$(\varphi \to \psi) \vee (\psi \to \varphi)\,;$$

(c) the product logic Π *if one adds the axiom schemata*

$$\varphi \wedge \psi \to \varphi \,\&\, (\varphi \to \chi)\,,$$
$$(\varphi \to \psi) \vee (\psi \to \varphi)\,,$$
$$\varphi \wedge \psi \to \bar{0}\,,$$
$$\neg\neg\chi \to ((\varphi \,\&\, \chi \to \psi \,\&\, \chi) \to (\varphi \to \psi))\,.$$

This monoidal logic covers also further systems of non-classical logic. The most important is GIRARD's *commutative linear logic* which may be characterized, cf. [7], by the class of all those residuated lattices which satisfy the law: $a = (a \rightarrowtail 0) \rightarrowtail 0$.

Proposition 7 *An adequate axiomatization of the commutative linear logic of* GIRARD *is given by the axiom system* $\mathsf{Ax_{ML}}$ *together with the additional schema:* $\neg\neg\varphi \to \varphi$.

And also classical (propositional) logic can be axiomatized by an extension of $\mathsf{Ax_{ML}}$.

Proposition 8 *An adequate axiomatization of classical propositional logic is given by the axiom system* $\mathsf{Ax_{ML}}$ *together with the additional schemata*

$$\varphi \wedge \psi \to \varphi \,\&\, (\varphi \to \psi)\,,$$
$$\neg\neg\varphi \to \varphi\,,$$
$$\varphi \to \varphi \,\&\, \varphi\,.$$

2.2 Basic logic

The axiomatization of HÁJEK [3] for the basic logic, i.e. for the class of all wffs which are valid in all BL-algebras, is given in a language which has as basic vocabulary the connectives $\to, \&$ and the truth degree constant $\bar{0}$, taken in each BL-algebra $\langle L, \cap, \cup, *, \rightarrowtail, 0, 1\rangle$ as the operations $\rightarrowtail, *$ and the element 0. Then basic logic has as axiom system $\mathsf{Ax_{BL}}$ the following schemata:

$$(\varphi \to \psi) \to ((\psi \to \chi) \to (\varphi \to \chi))\,,$$
$$\varphi \,\&\, \psi \to \varphi\,,$$
$$\varphi \,\&\, \psi \to \psi \,\&\, \varphi\,,$$
$$\varphi \,\&\, (\varphi \to \psi) \to \psi \,\&\, (\chi \to \varphi)\,,$$
$$(\varphi \to (\psi \to \chi)) \to (\varphi \,\&\, \psi \to \chi)\,,$$
$$(\varphi \,\&\, \psi \to \chi) \to (\varphi \to (\psi \to \chi))\,,$$
$$((\varphi \to \psi) \to \chi) \to$$
$$\quad (((\chi \to \varphi) \to \chi) \to \chi)\,,$$
$$\bar{0} \to \varphi\,,$$

and has as its (only) inference rule the rule of detachment (w.r.t. the implication connective $\to$).

The logical calculus which is constituted by this axiom system and its inference rule, and which has the standard notion of derivation, shall be denoted by $\mathbb{K}_{\mathsf{BL}}$.

Usually the language is extended by definitions of additional connectives and a further truth degree constant:

$$\varphi \wedge \psi \;=_{\text{def}}\; \varphi \,\&\, (\varphi \to \psi)\,,$$
$$\varphi \vee \psi \;=_{\text{def}}\; ((\varphi \to \chi) \to \psi) \wedge ((\chi \to \varphi) \to \varphi)\,,$$

together with $\bar{1} =_{\text{def}} \bar{0} \to \bar{0}$.

Calculations (in BL-algebras) show that the additional connectives $\wedge, \vee$ just have the BL-algebraic operations $\cap, \cup$ as their truth degree functions.

It is a routine matter, but a bit tedious, to check that this logical calculus $\mathbb{K}_{\mathsf{BL}}$ is sound, i.e. derives only such formulas which are valid in all BL-algebras.

Corollary 9 *The axioms of the monoidal logic are* $\mathbb{K}_{\mathsf{BL}}$*-derivable.*

Corollary 10 *The* LINDENBAUM *algebra of the logical calculus* $\mathbb{K}_{\mathsf{BL}}$ *is a BL-algebra.*

Theorem 11 (Completeness) *For a wff* H *of* $\mathcal{L}_t$ *there are equivalent:*
(i) H *is derivable within the logical calculus* $\mathbb{K}_{\mathsf{BL}}$*;*
(ii) H *is valid in all BL-algebras;*
(iii) H *is valid in all linearly ordered BL-algebras, i.e. in all BL-chains.*

Theorem 12 *From the basic logic one gets adequate axiomatizations of:*
(a) the system L_∞ *if one adds the axiom schema:*

$$\neg\neg\varphi \to \varphi\,;$$

(b) the system G_∞ *if one adds the axiom schema:*

$$\varphi \to \varphi\,\&\,\varphi\,;$$

(c) the product logic Π *if one adds the axiom schemata*

$$\varphi \wedge \psi \to \bar{0}\,,$$
$$\neg\neg\chi \to$$
$$((\varphi\,\&\,\chi \to \psi\,\&\,\chi) \to (\varphi \to \psi))\,.$$

2.3 The logic of continuous t-norms

From the preceding considerations it is clear that all the logically valid wffs of monoidal (resp. basic) logic are valid in all t-norm based residuated lattices with a left continuous (resp. a continuous) t-norm.

Problem thus are whether there exist wffs (i) which are valid in all t-norm based residuated lattices with a left continuous t-norm but are not theorems of monoidal logic, and (ii) which are valid in all t-norm based residuated lattices with a continuous t-norm but are not theorems of basic logic.

The first problem is actually an open one. But the second problem has a negative answer, given by the following theorem.

Theorem 13 (Completeness) *The class of all wff which are provable in basic logic coincides with the class of all wffs which are logically valid in all t-norm based residuated lattices with a continuous t-norm.*

The main steps in the proof are to show (i) that each BL-algebra is a subdirect product of BL-chains, i.e. of linearly ordered BL-algebras, and (ii) that each BL-chain is the ordinal sum of BL-chains which are either trivial one-element BL-chains, or linearly ordered MV-algebras, or linearly ordered product algebras (as introduced in [5]), such that (iii) each such ordinal summand is locally embeddable into a t-norm based residuated lattices with a continuous t-norm, cf. [1, 4].

References

[1] CIGNOLI, R. - ESTEVA, F. - GODO, L. - TORRENS, A. (1999?). Basic Fuzzy Logic is the logic of continuous t-norms and their residua, *Soft Computing* (to appear).

[2] GOTTWALD, S. (1999?). *Many-Valued Logic.* Main Systems and Applications. (book in preparation)

[3] HÁJEK, P. (1998). *Metamathematics of Fuzzy Logic.* Kluwer Acad. Publ., Dordrecht.

[4] HÁJEK, P. (1998). Basic fuzzy logic and BL-algebras, *Soft Computing* **2**, 124–128.

[5] HÁJEK, P. – GODO, L. – ESTEVA, F. (1996). A complete many-valued logic with product-conjunction. *Arch. Math. Logic* **35**, 191–208.

[6] HÖHLE, U. (1994). Monoidal logic, in: Kruse/Gebhardt/Palm (eds.), *Fuzzy Systems in Computer Science,* Vieweg, Wiesbaden, 233–243.

[7] HÖHLE, U. (1996). On the fundamentals of fuzzy set theory; *J. Math. Anal. Appl.* **201**, 786–826.

HIGH PERFORMANCE FUZZY ASIC

L. Salvador P. Bernad D. Madroño
Instituto Nacional de Técnica Aeroespacial
Lab. de Hardware y Control
Ctra. Ajalvir km. 4 – 28850 – Torrejón de Ardoz – Madrid
salvadorla@inta.es

Abstract

In this paper we will describe the design and implementation of prototypes of fuzzy microprocessors over FPGA and ASIC.

The aim of this project is to achieve the implementation of efficient fuzzy hardware processors well suited for industrial applications. Another purpose is to get high technical skills in the synthesis of high performance digital systems for information processing.

The first purpose is accomplished by the design of a three pipelined module processor that involves systolic arrays on each stage to reach high performance. The second one refers to the synthesis of several prototypes over FPGA (ACTEL, XILINX) previous to send the final system to be manufactured in an fifty thousand gates ASIC. Part of this prototypes are embedded in PCB that are able to communicate via Labview acquisition boards to a host and play as coprocessor or accelerators for fuzzy applications.

Parallel to those developments, we have carried out some research to obtain more effective implementations of operators like Lukasiewicz or Max-Prod [10], [7].

Keywords: Hardware, fuzzy, FPGA, ASIC, microprocessor, Lukasiewicz.

1 PROCESSOR SPECIFICATIONS

The main purpose of this project is the implementation of a fuzzy microprocessor of high performance. Performance parameter is not only the inference speed but the complexity of fuzzy model supported and the flexibility of configuration and operation tasks of the digital circuit.

The system must compute over five millions of fuzzy inferences per second of the whole rule base (not only a rule). It must support fuzzy variables and crisp ones. The rules must be in complex form, therefore, they must include the operators AND, OR and NOT [1] and fuzzy modifiers [2]. The number of rules must be at least thirty and support several context of operations [3].

The configuration must be performed in real time and the circuits must be able to operate in burst mode or single mode. Moreover, each one of the modules of the circuit (fuzzifier, inference and defuzzifier) can be manage independently.

2 DESIGN FLOW

In this section we will describe the way the project has been structured to achieve the objectives described above.

2.1 Processor Structure

The processor is structured in four main parts:

- Processor environment
- Fuzzifier
- Defuzzifier
- Inference.

The last three parts are called the Core of the processor and it is the classic division of a fuzzy processor. The relations between those modules in this implementation is depicted in Figure 1.

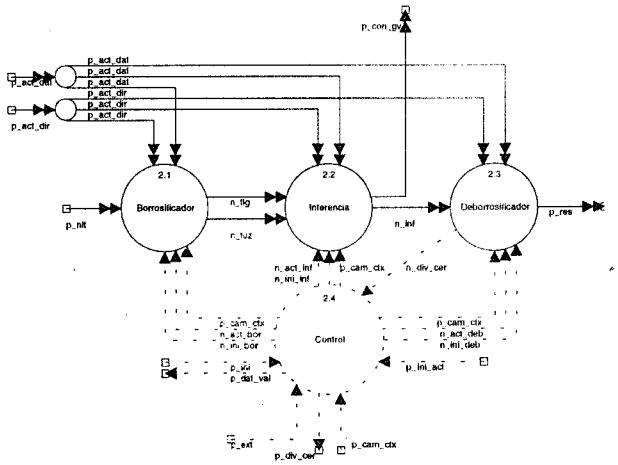


Figure 1 Relation between Core modules

The processor environment is the set of circuits that connects the general interface of the Core modules to a specific one.

The design of all the modules of the processor is based on bit-serial and digit-serial structures over pipelined and systolic architectures [4], [5], [6], [8], [9], [12].

Structures of processor have been designed as replication of elemental modules. Using this method system is easily scaleable for different fuzzy models.

The full microprocessor is configured before operation using a common bus of data and addresses. This design will support the interface with a standard bus as PCI or VME in the future.

2.2 Development Flow

The development of the microprocessor has been based on an incremental flow, as describes **Figure 2**.

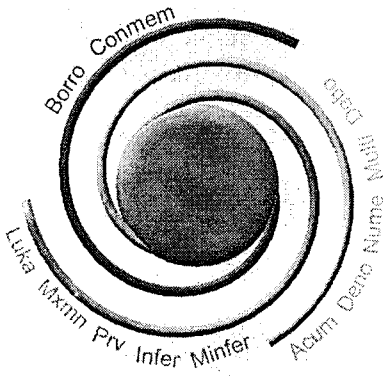


Figure 2 Development flow

The develop of the final microprocessor has been made through the design and implementation of prototypes more and more complex. The final result of the evolution of these prototypes are the four modules enumerated above. The integration of the four ones is the final processor.

Every part of microprocessor has been specified, described in VHDL, simulated, synthesised, built and probed physically, as describes the technical flow.

2.3 Technical Flow

The employed technical flow for each prototype is based on VHDL design. It has the following stages:

1) VHDL design and integration of modules
2) FPGA synthesis
3) Prototype tests
4) PCB building
5) Operation environment

The first three stages are described in **Figure 3**. This image employs the icons of different applications used along the design process (but not all ones).

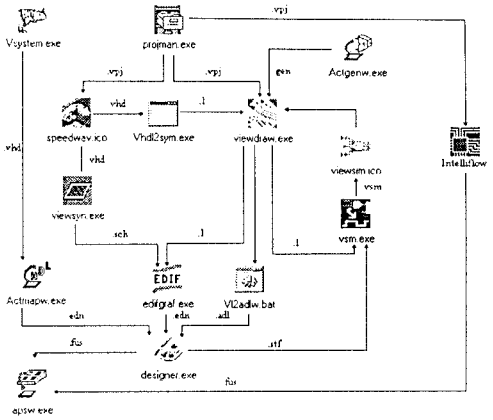


Figure 3 First three stages of technical Flow

Almost all prototypes are embedded in a PCB board developed in our laboratory. This board includes the connectors and circuits needed to interface with the PC-host. An example of these implementation is shown in **Figure 4**.

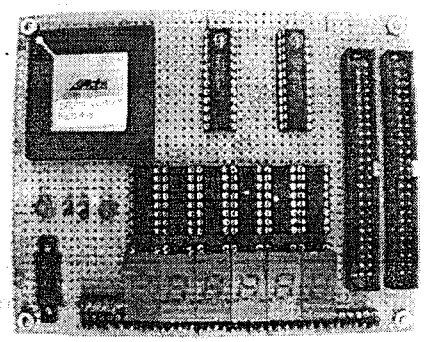


Figure 4 PCB of Deno prototype

The operational environment is formed by a host, a National Instrument acquisition board PCI-DIO/[illegible] the control software programmed on Labview. The whole system is shown in **Figure 5**.

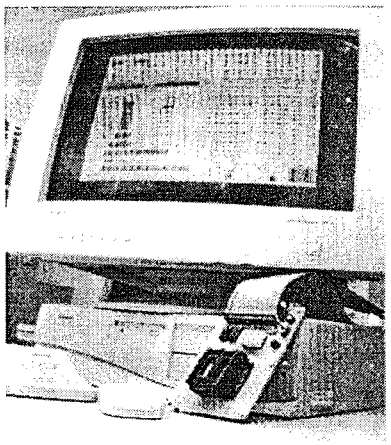


Figure 5 Operational environment

In the previous figure we can see one of the prototypes (right bottom side) connected via ribbon cable to the National DIO-96 (CPU back) and the software of operation. The software interface allows to control the configuration and operation of the prototype. Each software interface is specific for the device under test or operation. It has been programmed employing Labview. A detail of one of this interface is shown in **Figure 6**.

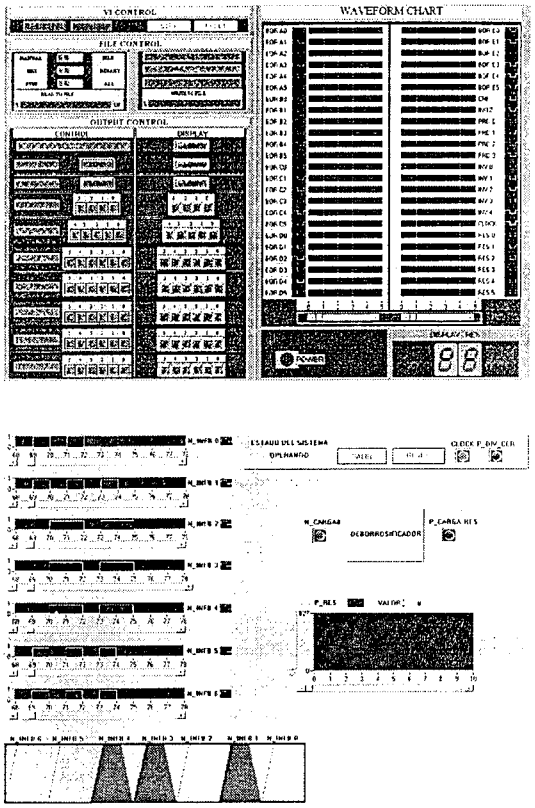


Figure 6 Details of software interface

The last interfaces has been developed adapting a common software interface that support the same format of input files than simulation environment and reduce the development time and fails rate.

3 PROTOTYPES AND PARTIAL RESULTS

The prototypes are implemented in FPGA's of ACTEL and XILINX families. The name of the prototypes are:

- Inference: Luka, Mxmn, Prv, Infer, Minfer.
- Defuzzifier: Acum, Deno, Nume, Multi, Debo
- Fuzzifier: Conmem, Borro
- Environment: Select, Select2m5

Several versions were made for each prototype. The performance of every one depends on the technology implemented and the speed grade of every particular chip. It comes from 10 MHz for the Acum prototype implementation on ACT1 family to the 50 MHz for the Prv prototype implementation on the 1200XL family speed grade –2.

The following collage shows physically the implementation on PCB of some of this prototypes:

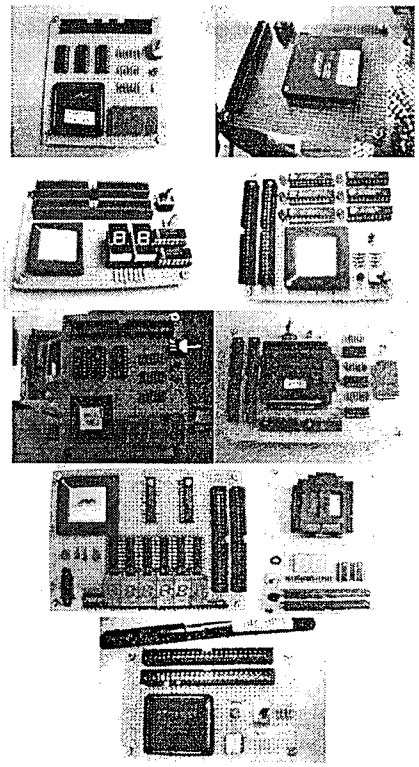


Figure 7 Some of the PCB prototype implementations

Development time for each prototype is has been about half one month. This time include all the design, build and test cycle.

4 FINAL SYSTEM

The final microprocessor is called MICRO00. It integrates the prototypes described in the previous section and it has been implemented twice:

- An ASIC version.
- An FPGA version.

The ASIC version is implemented using the Chip-Express libraries in technology of 0.35 µm. The clock speed is over 60 MHz and the encapsulation is a 210 pins plastic QFP.

This speed allows ten millions of evaluations per second of the complete rules database.

ASIC implementation is oriented to high performance environments. The manufacturing cost of each piece is still very high to be employed in home or educational applications.

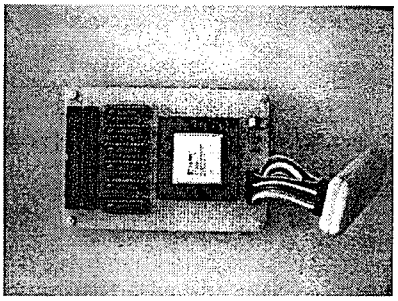


Figure 6 Xilinx implementation of microprocessor

The Xilinx implementation employs a XC4085 in a 559 pins plastic PGA. The clock speed is over 25 MHz.

The Xilinx XC4085 is programmed every time that system is on. It has a lot of free space yet. In the future it will can implement other models of fuzzy processing using this board and developed models.

Both implementations accomplish the requirements described at the beginning of this paper. They have been embedded in PCB boards to interface with a host and operational software environment developed in Labview.

Acknowledgements

This project is granted by the Spanish Government CICYT TIC-96-1393-C06-2 and it is a part of a main project named "Fuzzy microprocessor and its development environment". The last one involve several institutions and universities in Spain.

References

[1] E. Trillas, "Sobre funciones de negación en la teoría de conjuntos difusos", Stochastica, vol. 3, n° 1, 1979

[2] J. Gutiérrez., Luis de Salvador, "The Use of Modifiers in Fuzzy Systems and their Hardware Implementation", Congreso Internacional FUZZ-IEEE'97 Barcelona, 1997

[3] J.L. Castro, J.M. Zurita.(1992). "Controlador Difuso basado en Contextos" II Congreso Español sobre Tecnologías y Lógica Fuzzy.

[4] K.K. Parhi, "A systematic approach for design of digit-serial signal processing architectures", IEEE Transactions on Circuits and Systems, vol. 38, n° 4, abril 1991

[5] L. de Salvador, J. Gutiérrez, "A multilevel systolic approach for fuzzy inference hardware", IEEE Micro Magazine, Octubre 1995

[6] L. de Salvador, J. Gutiérrez, "High perfomance multilevel systolic fuzzy inference unit", Proceedings del I Congreso Conjunto NAFIPS/IFIS/NASA, 1994

[7] L. Salvador, P. Bernad, J. Gutiérrez "Circuito Experimental de Inferencia de Lukasiewicz", ESTYLF'98

[8] M.A. Manzoul, "Fuzzy controllers on semi-custom VLSI chips", Fuzzy Control Systems, Edt. A. Kandel, G. Laugholz, CRC Press, 1994

[9] Mahmoud A. Manzoul, S. Tayal.(1990) Sistolic VLSI array of multi-variable fuzzy control systems. Cybernetics and Systems: An International Journal 21. Pags 27-42.

[10] O. Cordón, F. Herrera, A. Peregrín, "Applicability of the fuzzy operators in the design of fuzzy logic controllers", Informe Técnico DECSAI-95111, Universidad de Granada, 1995

[11] P. Bernad, "Entorno multicontexto y metacontrol en microprocesadores borrosos", ESTYLF'97 p-263-269

[12] R.I. Hartley, K.K. Parhi, Digit-Serial Computation, Kluwer Academic Publishers, 1995

Lógica fuzzy y automatización: ¿Perversión lógica o revolución informática?

(A través de McNeil-Freiberger, Kahlert-Frank, Haack y Sangalli)

CARLOS ORTIZ DE LANDAZURI

Universidad de Navarra.

Palabras clave: indeterminado, borroso, nomonotónico, probable.

Presentación:

En 1965 Lotfi Zadeh propuso un procedimiento para evitar los *círculos viciosos* en sí mismos paradójicos, que habían provocado la *crisis de fundamentos de la teoría de conjuntos*. Dada la imposibilidad de aportar un procedimiento axiomático capaz de evitarlos, se optó por intentar neutralizar esta situación por un recurso de tipo práctico, como fue la formalización de los conjuntos *borrosos* o *imprecisos*, a los que se atribuía el ser los causantes de esta crisis[1]. Sin embargo fueron numerosas las críticas que desde un principio acompañaron la génesis y el posterior desarrollo de este tipo de lógica: se definió como una *lógica imprecisa*, *borrosa*, o simplemente *mal fundada*, dando un claro sentido despectivo a estos términos. Sin embargo el paso de tiempo demostró que las propuestas de Zadeh presentaban virtualidades muy novedosas, especialmente en el campo de la informática y de la automatización[2]. Pero tampoco le han faltado críticas por su pretensión de proponerse como una 'lógica alternativa' capaz de lograr de este modo una posible *fundamentación de la teoría de conjuntos*, como al menos ha sugerido Susan Haack[3].

1.- Zadeh, 1965-1995: La vuelta a una lógica fuzzy 30 años después.

Susan Haack en 1976 y 1996 ha revisado los planteamientos de Lotfi Zadeh en 1965, cuando propuso una ampliación de la noción de conjunto o clase, a partir de la elaboración de una nueva *lógica fuzzy*, con unas pretensiones desproporcionadas[4]. Según Susan Haack, Zadeh, propuso una reformulación de la teoría de conjuntos, a partir de una nueva redefinición de sus *elementos originarios* sin excluir ninguna noción básica, como son las propias nociones de conjunto, o de verdad, o de validez. De este modo llevó a cabo un análisis de determinados presupuestos teóricos que, sin duda alguna, han afectado a la fundamentación de la teoría de conjuntos, a pesar de la separación que hoy por hoy existe entre ambas disciplinas[5].

2.- El programa inicial de Zadeh en 1965.

En estos casos la *lógica fuzzy* tomó como punto de partida conjuntos imprecisos o difusos, como el ser más alto, o de mayor edad, en sí mismos relativos. Estos conjuntos no cumplen con algunas de las condiciones exigidas por el proyecto *"New Fundations"* para los *conjuntos bien fundados*, al menos según algunos seguidores de Quine[6], o más bien Peirce, como ahora sugiere Haack, revisando en parte sus anteriones propuestas. A este respecto la lógica fuzzy aplica a este tipo de conjuntos una *lógica no-bivalente* cuyos elementos originarios pueden adolecer de una validez veritativa en sí misma indeterminada, sin ver en ello un defecto formal insubsanable. Por ello estos elementos originarios tampoco se tipifican como conjuntos *no bien fundados*, o simplemente *antifundados*, al modo como sucedía anteriormente en el proyecto del *'New Fundation'*[7]. Más bien se toman como el punto de partida originario de la filosofía del conjuntos sin introducir condiciones complementarias que, en su opinión, vienen impuestas por un formalismo externo al uso habitual de los conceptos o clases[8].

A este respecto la lógica fuzzy no justifica su validez en virtud de una lógica de *primer orden* de tipo *monotónica*, donde el valor de verdad de una proposición permanece fijo para una serie infinita de elementos, como en cambio sucede en la teoría de conjuntos tradicional. En su lugar los conectivos lógicos, incluido el condicional, también admiten el posible cambio del grado de verdad de una proposición, sin darle la interpretación unívoca de tipo monotónico, como ha sido habitual[9]. En este sentido los conjuntos borrosos son conjuntos *finitos*, que desde un punto de vista formal no están bien ordenados, ni definen con precisión el lugar que cada uno de los elementos ocupan respecto de todo el conjunto, o al menos admiten la posibilidad de no estarlo[10]. Por todo ello la lógica fuzzy adopta una actitud inicialmente abierta ante las distintas formas de predicación, ya sean de tipo nomonotónico, o simplemente analógica, siempre que se refieran a series finitas previamente delimitadas de un modo protocolario, aunque evidentemente este tipo de propuestas también han levantado polémica. A este respecto he tomado cuatro ejemplos de como hoy día se abordan estos problemas de fundamentación última por parte de la *lógica fuzzy*, sin ánimo de ser exhaustivo: Dos a favor, uno en contra y, finalmente, otro que intenta mediar entre ambos[11].

3.- Daniel McNeil y Paul Freiberger (1993): Las virtualidades metodológicas e informáticas de la lógica fuzzy.

Daniel McNeil y Paul Freiberger[12] han aplicado la *lógica fuzzy* al análisis de algunas nociones metodológicas básicas de la *filosofía de la ciencia y de la técnica*, con un

---

[1] Cf. Blum, W.; *Die Grammatik der Logik. Einführung in die Mathematik*, Deutscher Taschenbuch, Müchen, 1999.

[2] Cf. Wilson, M. D.; *Ideas and Mechanism. Essays on Early Modern Philosophy*, Princeton, New Jersey, 1999.

[3] Cf. Haack, S.; *Deviant Logic, Fuzzy Logic.Beyond the Formalism*, The University of Chicago Press, Chicago, 1974, 1996.

[4] Cf. Zimmermann, H-J.; *Fuzzy set theory and its applications*, Kluwer, Boston, 1991.

[5] Cf. Gupta, M. M.; Yamakawa, T. (ed); *Fuzzy computing: theory, hardware and applications*, North Holland, Amsterdam, 1988.

[6] Cf. Hahn, L.; Schilpp, P. A. (ed); *The Philosophy of W. V. Quine*, Open Court, London, 1997.

[7] Cf. Barwise, J.; Moss. L.; *Vicious Circles. On the Mathematic of Non-Wellfounded Phenomena*, CSLI, Standford University, Standford, 1996.

[8] Cf. Kosko, B.; *Fuzzy Thinking: The New Science of Fuzzy Logic*, Hyperion, New York, 1993.

[9] Cf. Ruppen, P.: *Einstieg in die formale Logik*, Peter Lang, Bern, 1997.

[10] Cf. AA. VV.; *Fuzzy Sets and Systems. International Journal of Soft Computing and Intelligence. Vol 94, Nº1, 1998*, North-Holland, Amsterdam.

[11] Cf. Kaufmann, A.; Gupta, M. M. (ed); *Fuzzy mathematical models in engineering and management science*, North Holland, Amsterdam, 1991.

[12] Cf. McNeil, D.; Freiberger, P.; *Fuzzy Logic*, Simon & Schuster, New York, 1993; *Fuzzy Logic. Die 'unscharfe' Logik erobert die Technik*, Droemer Knaur, München, 1994.

estilo divulgador fácilmente comprensible. En su opinión, la *lógica fuzzy* permite proseguir la fundamentación de la teoría de la ciencia hasta niveles anteriormente inaccesibles[13]. En este sentido el recurso a conceptos imprecisos requiere el uso compartido de un instrumental analítico paradójicamente muy preciso, que a su vez permite articular nociones sutiles de difícil manejo. Por ejemplo, los múltiples sentidos que la noción de *probabilidad* ha adquirido en los últimos años a través del propio desarrollo científico y técnico, produciendo un vuelco en el modo de abordar unos problemas que hace poco eran irresolubles[14].

En su opinión, la *lógica fuzzy* no es una lógica imprecisa, ni tampoco de éxito seguro. Simplemente ofrece un poderosos instrumental lógico con virtualidades múltiples que hay que saber extraer en cada caso; especialmente interesante es su propuesta de lograr una posible integración entre los distintos usos de la probabilidad, sin que hasta ahora se haya podido establecer un puente de unión entre ella. Por un lado la probabilidad subjetiva o inversa, siguiendo el método de Bayes; y, por otro lado, la probabilidad objetiva o frecuencial, de uso más habitual. Con este fin se aplica a la noción de probalidad los grados de verdad característicos de la *lógica fuzzy,* según un criterio de máximos y mínimos. Se pone de manifiesto así como las dos nociones citadas de probabilidad sólo expresan dos situaciones límites de mínima y máxima información, según se conciba la probabilidad desde un punto de vista finitista o, por el contrario, desde un punto de vista infinitista, o estrictamente frecuencial. Posteriormente se aplica este mismo análisis propio de los *sistemas expertos* a otras muchas nociones básicas de la lógica de la ciencia y de la teoría de conjuntos, como ya demostró Reichenbach respecto de la física en general[15].

4.- Kahlert y Frank (1994): Hacia una fundamentación de la teoría de conjuntos a partir de una logica fuzzy.

Jörg Kahlert y Hubert Frank[16] en 1994 han dado un paso más aplicando la *lógica fuzzy* al análisis de algunas nociones básicas de *los sistemas expertos*. En su opinión, los sistemas de control propuestos por la *lógica fuzzy* permiten proseguir el análisis de algunos conceptos o nociones originarias, como son las propias nociones de *entroncamiento* (*"entailment"*) o de *concatenación* lógica necesaria (*"Verknüpfungen"*), más allá de donde estas nociones podían ser analizadas por la teoría de conjuntos, o incluso por la propia lógica modal clásica. Con este fin se propone un nuevo modo de conceptualizar los grados de verdad como si se tratara de un sistema no bivalente, con máximos y mínimos, al modo de la lógica inductiva de Peirce[17].

La *lógica fuzzy* retrotrae así el proceso de fundamentación de la teoría de conjuntos a un momento previo, en donde también se tienen en cuenta los conjuntos *no-bien-fundados*, o simplemente *antifundados*, al igual que ocurre en el proyecto *"New Fundation"* , y en general en la lógica booleana de segundo orden[18]. Sin embargo Lotfi Zadeh no pretendió tanto, y nunca justificó esta posible relación de subordinación interna entre los conjuntos que están bien fundados y los que no[19]. En cambio para Kahlert y Frank, la lógica fuzzy localiza una fase previa de fundamentación, que hasta hace muy poco no había sido tenido en cuenta por la teoría clásica de conjuntos, aunque de hecho sólo se fija casi exclusivamente en sus posteriores aplicaciones prácticas[20]. Aporta un nuevo modo de conceptualizar las *reglas lógicas* básicas de la inclusión, la conjunción y del condicional. Amplía los modos posibles de formalizar el uso de los *primeros principios* de la lógica, con esquemas propios de inferencia. Articula las diversas formas de *probabilidad*, y aporta un fundamento algorítmico para los *sistemas expertos*[21].

5.- Susan Haack,1974-1996: El fundamento perverso de la lógica fuzzy.

Susan Haack en 1996, en la reedición de *Logica pervertida, Logica Fuzzy. Más allá del formalismo* [22], reconoce los enormes progresos que la *lógica fuzzy* ha experimentado hoy día, con aplicaciones espectaculares especialmente en el ámbito tecnológico, pero reafirma en sus anteriores valoraciones respecto a los planteamientos iniciales de Lotfi Zadeh. Se congratula de este éxito inesperado, aunque sigue sin compartir las motivaciones más profundas que llevaron a Zadeh y después a Bart Kosko[23] a hacer este tipo de propuestas, con pretensiones en su opinión desproporcionadas. En su opinión, Zadeh concibió la *lógica fuzzy* como una teoría rival que se presenta como una alternativa efectiva a la teoría de conjuntos clásica. La *lógica fuzzy* no sólo pretende una ampliación de la lógica a un campo nuevo, como son los conjuntos o clases borrosas de definición imprecisa, sino que su propuesta va más allá: reivindica un nuevo modo de conceptualizar y formalizar los conceptos básicos de la lógica, como son la noción de verdad, de probabilidad[24], o de validez. Pero según Haack, esta lógica siempre mantiene una dependencia respecto del modo clásico como Boole, Peirce o Frege abordaron estos problemas de fundamentación última, sin poderse presentar como una 'lógica alternativa' verdaderamente solvente[25].

Según Susan Haack, cuando la *lógica fuzzy* se independiza de la lógica clásica, entonces relativiza los elementos originarios claves de la teoría de conjuntos, sin poder garantizar su propia validez[26]. Por ejemplo cuando Zadeh afirma *que 'la verdad es un predicado fuzzy'*, o que la *'lógica se remite a valores y grados de verdad igualmente fuzzys'*[27]. En estos casos la noción básica de *verdad* y *falsedad*; se conceptualiza como un conjunto de elementos no correctamente ordenados, o no bien fundados, a pesar de afirmarse como un presupuesto implícito de la propia

---

[13] Cf. Hopkins, R.; *Picture, Image and Experience. A Philosophical Inquiry*, Cambridge University, Cambridge, 1998.

[14] Cf. Cohen, L. J.; *Introduzione alla filosofia dell'induzione e della probabilitá*, Giuffrè, Milano, 1999.

[15] Cf. Poser, H.; Dirks, U (Hrsg); *Hans Reinchenbach. Philosophie im Umkreis der Physics*, Akademic, Berlin, 1999.

[16] Cf. Kahlert, J.; Frank, H.; *Fuzzy-Logik und Fuzzy-control. Eine anwendungsorientierte Einbführung mit Begleitsoftware*, Vieweg & Sohn, Braunschweig/Wiesbaden, 1994.

[17] Cf. Kuhn, F.; *Ein anderes Bild des Pragmatismus*, Klostermann, Frankfurt, 1996.

[18] Cf. Cf. Hughes, G. E.; Cresswell, M. J.; *A New Introduction to Modal Logic,* Routledge, 1996.

[19] Cf. Kaufmann, A.; *Introduction to fuzzy theory and applications.* Van Nostrand Reinhold, New York, 1985.

[20] Cf. . Howson, C.: *Logic with Trees. An Introduction to Symbolic Logic,* Routledge, London, 1997.

[21] Cf. Adams, E. W.; *A primer of Probability Logic*, CSLI, Standford University, California, 1998.

[22] Haack, S.; Ibidem. Cf. Kahlc, W.; Collani, E. von; Franz, J.; Jensen, U.; *Advances in Stochastic Models for Reliability, Quality and Safety*, Birkhauser, Basel, 1998.

[23] Cf. Haack, S. p. 230. Kosko, B.; *Fuzzy Thinking: The New Science of Fuzzy Logic*, Hyperion, New York, 1993.

[24] Cf. Logue, J.: *Projective Probability*, Clarendon, Oxford, 1995.

[25] Cf. Dorn, G. J. W.; *Deductive, Probabilistic and Inductive Dependence. An Axiomatic Study in Probability Semantics*, Peter Lang, Frankfurt, 1997.

[26] Cf. Gasarch, W. I.; Martin, G. A.; *Bounded Queries in Recursion Theory*, Birkhäuser, Boston, 1999.

[27] Cf. Zadeh, L.; 'Fuzzy Sets'. *Information and Control* 8: 338-53, 1965; 'Fuzzy Logic and Approximate Reasoning', Synthese 30: 407-25, 1975.

formalización de la teoría de conjuntos. Susan Haack no rechaza las posibles aportaciones que pueden hacer las así llamadas lógicas *"alternativas"* respecto de una posible renovación de los planteamientos clásicos. Ella misma hace notar las expectativas que generó Peirce al proponer una *lógica de la imprecisión o vaguedad* ("Vaguenes")[28] o una lógica no-bivalente, que admite grados intermedios entre la verdad y la falsedad, a través por ejemplo de su propia noción de probabilidad subjetiva, o inductiva, sin que en ningún momento este tipo de propuestas haya supuesto una seria amenaza para los planteamientos clásicos. Sin embargo rechaza que la *lógica fuzzy*, al igual que otras así llamadas *"lógicas alternativas"* puedan aportar algo a la resolución de las paradojas que aún hoy día genera la teoría de conjuntos, u otras nociones metodológicas básicas, como la propia noción de probabilidad. En todo caso antes debería definir con mas claridad su postura respecto a otras nociones aún más elementales, como las ahora señaladas[29].

En su opinión, la *lógica fuzzy* en esos casos pierde la sensibilidad filosófica necesaria para plantearse este tipo de problemas con una mínima *"bona fides"* filosófica. Por ejemplo, se admite la existencia de posibles grados de verdad con un creciente relativismo, sin poder garantizar tampoco la posible validez de sus propias reglas de argumentación[30]. Todo ello le impide ser una *lógica estricta*, cosa que al parecer tampoco les preocupa. En este sentido Susan Haack califica a la *lógica fuzzy* de *lógica pervertida*, al menos cuando se la concibe como una disciplina rival a la teoría clásica de conjuntos de Boole, Frege o Peirce, como propuso Zadeh. Sin embargo no es el único caso. Según Haack, algo similar sucede con la *lógica objetos inexistentes* de Meinong, o con la así llamada *lógica paraconsistente de algunos dialécticos*, o con la *lógica intuicionista* de Dummett, o con las lógicas recíprocamente *inconmensurables* de Quine[31].

6.- Sangalli, 1998: Las virtualidades teóricas y prácticas de la lógica fuzzy.

Arturo Sangalli ha defendido en *"La importancia de ser fuzzy"*[32], los planteamientos de Lotfi Zadeh en 1965. En su opinión, los 43 años transcurridos no han hecho sino confirmar las múltiples virtualidades teóricas y prácticas de sus propuestas. A su modo de ver la lógica fuzzy adolece de un conjunto de problemas debido al mal uso que se hicieron de las propuestas de Lotfi Zadeh, buscando aplicaciones en el cálculo de probabilidades[33], o en el ámbito estrictamente metodológico, o en la fundamentación de las matemáticas, cuando sus propuestas se referían a un ámbito muy concreto y delimitado[34]: lograr una formalización de los *conjuntos imprecisos* o *borrosos*, mediante una ampliación de la lógica booleana de la teoría de conjuntos bien formados, y la aplicación de otras técnicas complementarias; por ejemplo, en 1965 Zadeh usó ciertos gráficos para representar el tipo de imprecisión de un conjunto borroso respecto de un modelo o arquetipo previamente fijado; o en 1990 Bart Kosko utilizó los ejes de coordenadas cartesianas para representar los *'grados de pertenencia'* de un determinado elemento a su correspondiente conjunto, desde un máximo hasta un mínimo, según cual fuera el grado de imprecisión o borrosidad; o en 1991 Hitachi ideó un dispositivo para identificar la autoria de una firma, en razón de los distintos grados de fiabilidad de la gráfica, como son los trazos, o la velocidad de escritura; en estos casos, a medida que *crece la complejidad* de un análisis se incrementa a su vez el *grado de incertidumbre* de sus conclusiones, aunque no por ello se renuncia a la posibilidad de enjuiciar el *grado de pertenencia* que en cada caso le corresponde respecto de un modelo previamente fijado como término de referencia[35].

Evidentemente estas propuestas siempre presuponen la formulación del teorema de Parikh acerca de la definición de vaguedad esencial o inerradicable para los espacion α–conexos, así como la validez de las medidas de Sugeno respecto a los conjuntos fuzzy, sin confundir incerteza e imprecisión. De igual modo que se aceptó la formalización de la así llamada *fisura* de Herbrand, según se separen precisivamente las relaciones de tipo monotónico y nomonotómico, que generan los conjuntos borrosos, respecto al conjunto que genera la suma indistinta de ambos[36]. Pero, según Sangalli, no ocurre lo mismo con el cálculo de probabilidades, ya que se trata de fenómenos totalmente aleatorios donde no se puede prefijar de antemano su modo de comportamiento, ni el grado de azar, o de verdad, siendo fenómenos aleatorios totalmente incompatibles con este tipo de mediciones fuzzy. Simultaneamente este mismo tipo de lógica ha tenido múltiples aplicaciones en el ámbito de las redes neuronales[37], o en la programación de las computadoras, o en el desarrollo de la automatización y los sistemas expertos. El argumento decisivo a su favor ha sido la viabilidad práctica de sus distintas propuestas. Pero, según Sangalli, la lógica fuzzy nunca debe presentarse como una solución alternativa al problema de la fundamentación última de las matemáticas, sino que debe reconocer su dependencia interna respecto del cálculo booleano[38].

Según Arturo Sangalli, el cálculo booleano también se aplica a la lógica fuzzy. En su opinión, los conjuntos imprecisos o borrosos se pueden considerar clases bien formadas desde un punto de vista conceptual, aunque para nosotros aparentemente no lo sean debido a la complejidad de los factores analizados, o a la ausencia de una técnica adecuada para medir su peculiar tipo de precisión conceptual[39]. En este sentido la lógica fuzzy aporta técnicas muy precisas para abordar un tipo peculiar de objetos, al modo como también en la geometría clásica ocurrió con el método de las *exhauciones*, o de las aproximaciones al límite mediante sucesivas exclusiones; entonces también se reconoció la existencia de objetos que eran *inconmensurables* desde el punto de vista de su medición, pero a la vez conceptualmente muy precisos, como de hecho ocurrió con el valor de 'π'. Con este fin se idearon diversos sistemas alternativos de medición con la máxima precisión

---

[28] Cf. Keefe, R.; Smith, P. (ed); *Vagueness. A Reader*, MIT, Cambridge (MA), 1996. Smith, B. (ed); *Vagueness, The Monist*, April 1998, 81, 2, The Hegeler Institut, La Salle, Illinois.

[29] Cf. Suppes, P. Zanotti, M.: *Foundations of probability with applications,* Cambridge University Press, Cambridge, 1996.

[30] Cf. Arnauld, A./Nicole, P.: *Logic or the Art of Thinking*, Cambridge University Press, Cambridge, 1996.

[31] Cf. Read, S.: *Thinking about logic: an introduction to the philosophy of logic*, Oxford University Press, Oxford, 1994.

[32] Cf. Sangalli, A.; *The Importance of Being Fuzzy, And other Insights from the border between Math and Computers*, Princeton University, New Jersey, 1998, 208 págs.

[33] Cf. Dembski, W. A.; *The Design Inference. Eliminating Chance Through Small Probabilities*, Cambridge University, Cambridge, 1998.

[34] Cf. Schinazi, R. B.; *Classical and Spatial Stochastic Processes*, Birkhäuser, Boston, 1999.

[35] Cf, Resnick, S.; A Probability Path, Birkhäuser, Boston, 1999.

[36] Cf. Brewka, G.; Dix, J.; Konolige, K.; *Nonmonotonic Reasoning. An Overview*, CSLI, Stanford University, Stanford, 1997.

[37] Cf. Nauck, D.; Klawon, F.; Kruser. R. (ed); *Neuronale Netze und Fuzzy Systeme: Grundlagen des Konnektionismus, Neuronaler Fuzzy Systeme und koplung mitwissensbasierte Methoden*, Vieweg, Braunschweig, 1994.

[38] Cf. Watton, D.; *Argument Structures. A Pragmatic Theory*, University Press, Toronto, 1996.

[39] Cf. Schum, D. A.: *Evidential Foundations of Probabilistic Reasoning*, John Wiley & Sons, New York, 1994.

posible, aunque ahora presentarían una dificultad añadida: la *lógica fuzzy* también se refiere a conjuntos de múltiples variables, todas ellas igualmente imprecisas o borrosas desde el punto de vista de la medición, dando lugar a *hiperconjuntos* o *hipercubos* de infinitas dimensiones, sin que ello merme su capacidad de definición conceptual. En este sentido la lógica fuzzy exige pruebas, o *hiperpruebas metamatemáticas*, que garanticen la validez de sus proyectos, al modo como se ha hecho habitual en las distintas ramas de las matemáticas desde los tiempos de Hilbert en 1928, incluyendo incluso la posibilidad de una *automatización* de este tipo de cálculo, como ya propuso Turing en los años 30, sin ser tampoco en esto una excepción[40].

7.- Hacia una ampliación revolucionaria de la automatización informática.

Para justificar estas conclusiones la monografía se divide en tres partes, con seis capítulos en conjunto. En *'Visiones borrosas'* se analiza las dos características aparentemente opuestas de la lógica fuzzy: la incertidumbre de sus límites taxonómicos, y la precisión y efectividad de sus técnicas de control estratégico[41]. En *'Límites'* se describen las virtualidades teóricas y prácticas de la *lógica fuzzy*: por un lado, la ampliación de cálculo automático hasta ámbitos reacios a cualquier tipo de formalización; por ejemplo, los sistemas complejos mecánicamente indecidibles, pero capaces de valorar el grado de imprecisión o incertidumbre que en cada caso les corresponde[42]; por otro lado, la posible axiomatización y automatización de la *lógica fuzzy*, a fin de dar una respuesta a problemas hasta ahora considerados insolubles; por ejemplo, el así llamado mapa de los cuatro colores, resuelto definitivamente por Appel y Hanken en 1976 con ayuda de una computadora, mediante un método de exclusiones o exhauciones de los distintos gráficos mutuamente imcompatibles, que dan lugar a un cálculo matemático excesivamente complejo para la mente humana, pero automatizable. Finalmente, en *'Soluciones naturales'* se analizan dos ámbitos de aplicación de la lógica fuzzy: por un lado, el ámbito técnico-práctico de sus posibles aplicaciones a los modelos de redes neuronales, a la ingeniería genética, a los sistemas de seguridad informática, a la inteligencia artificial, o a la explotación agrícola[43]; por otro lado, el ámbito teórico-especulativo respecto de sus posibles aportaciones a la solución del problema de la selección natural, de los modelos evolutivos genéticos, del crecimiento de población o de los indices de supervivencia. Todo ello, finalmente, conduce a un replanteamiento del *dilema del prisionero*, en su nueva versión del *dilema del representante comercial*, donde toda opción resulta igualmente contraproducente, salvo que se recurra a unos 'criterios de pertenencia' o 'preferencia', como los ahora expuestos. A partir de aquí se propone una nueva valoración de la naturaleza del aprendizaje instrumental, o del coeficiente intelectual, o del propio sentido evolutivo de la historia natural[44].

Evidentemente Arturo Sangalli ha elaborado una tesis muy sugerente y ambiciosa que, sin duda alguna, dará lugar a posteriores desarrollos mas pormenorizados. A este respecto resulta de gran interés su afán por encontrar un punto de vista unificador, donde la lógica vuelve a desempeñar un papel regulador básico, al modo como también ocurre en las propuestas de Popper, o de Bertrand Russell[45]. Pero precisamente aquí reside la objeción más seria que se puede formular a su propuesta: ¿Realmente este programa de ampliación de la lógica fuzzy hasta los extremos que ahora se nos detalla es coherente con las propuestas de Lotfi Zadeh? ¿Realmente se puede considerar la lógica fuzzy como un mero complemento de la lógica booleana, manteniendo en todo momento una dependencia casi total respecto de sus propuestas teóricas, incluida ahora también la posible aplicación de esta lógica a la totalidad de la historia natural del universo, cuando Lotfi Zadeh, en opinión de Susan Haack, pretendió desarrollar una lógica alternativa totalmente abierta, sin prejuzgar por ello el orden existente en el universo físico, como en cambio ocurre en Boole[46].

8.- Conclusión: ¿Perversión lógica o condición humana?

Evidentemente hoy nadie discute las virtualidades tecnológicas de la *lógica fuzzy*, mientras que la teoría de conjuntos se reserva los problemas nunca suficientemente resueltos de la axiomatización y de la fundamentación última[47], sin que al parecer a la lógica fuzzy le preocupe excesivamente este reparto de papeles, ni la posible perversión de la lógica[48]. Sin embargo esta situación es al menos anómala y ambas disciplinas están llamadas a entenderse. A este respecto la *lógica fuzzy* antes o después tendrá que afrontar con sus propias técnicas de análisis estos problemas de fundamentación última que, al parecer, de momento ha ido soslayando; especialmente cuando se le acusa de parasitismo o dependencia de la lógica booleana, cuando a la vez ella misma reivindica una autonomía o carácter *alternativo* que, al menos para Susan Haack, sería sinónimo de una *perversión lógica*. Sin embargo la lógica fuzzy presenta a este respecto una novedad, que habitualmente no se tiene en cuenta: la *lógica fuzzy* hizo de la debilidad virtud y de la imprecisión su mejor defensa, poniendo de manifiesto como el rigor del análisis matemático es compatible con el reconocimiento de sus propias *condiciones de finitud*[49].

Carlos Ortiz de Landázuri
Pamplona - 12-IV-1999.

---

[40] Cf. Höhle, U.; *Mathematics of Fuzzy Sets*, Kluwer, Dordrecht, 1999.

[41] Cf. McEneaney, W. M.; Yin, G. G.; Zhang, Q. (eds); *Stochastic Analysis, Control, Optimization and Applications.A Volume in Honor of W. H. Fleming*, Birkhäuser, Boston, 1999.

[42] Cf. Weirich, P.; *Equilibrium and Rationality: Game Theory revisited by Decision Rules*, Cambridge University, Cambridge, 1998.

[43] Cf. Wirth, W.; Schweiger, W. (Hrsg.); *Selektion im Internet. Empirische Analysen zu einen Schlüsselkonzept*, Westdeutscher, Wiesbaden, 1999.

[44] Cf. Smithson, M.; *Fuzzy set analysis for behavioral and social sciences*, Springer, New York, 1987.

[45] Cf. Garcíadiego Dantan, A. R.; *Bertrand Russell and the Origins of the Set-theoretic 'paradoxes'*, Birkhauser, Basel, 1992.

[46] Cf. Nencka, H.; Bourguignon, J-P. (eds); *Geometry and Nature*, American Mathematical Society, Providence, 1997.

[47] Cf. Balakrishnan, N. (ed); *Advances in Combinatorial Methods and Applications to Probability and Statistics*, Birkhäuser, Boston, 1997, 562 págs.

[48] Cf. Kaplan, M.; *Decision Theory as Philosophy*, Cambridge University, Cambridge, 1996.

[49] Cf. Spangler, M.M. / Conway, P. (y otros): *Logic: an Aristotelian approach*, University Press of America, Lanham, 1993.

LÓGICA BORROSA E INTERPRETACIÓN CORRESPONDENTISTA DE LA TEORÍA SEMÁNTICA DE LA VERDAD DE TARSKI

María Teresa Casparri
e-mail: casparri@econ.uba.ar

Pablo S. García
e-mail: pablosgarcia@yahoo.com

Luisa L. Lazzari
e-mail: ilazzari@econ.uba.ar

Facultad de Ciencias Económicas - Universidad de Buenos Aires
Córdoba 2122, 2° piso, Buenos Aires (1120), Argentina.

Resumen

Hájek sostiene que la propiedad más importante de los predicados vagos es su carácter comparativo: el hecho de que puedan aplicarse a un objeto (individual) dado en mayor o menor medida, incluyendo el sí absoluto y el no absoluto. Esta propiedad comparativa de los predicados vagos nos conduce a una "noción comparativa de verdad", en virtud de la cual una proposición puede ser "más verdadera" o "menos verdadera", la misma proposición puede tener varios valores de verdad en diferentes situaciones: puede ser absolutamente verdadera, absolutamente falsa, o algo intermedio. Hájek vincula esta situación con la noción comparativa de existencia en los filósofos neoplatónicos, a la que considera, si no un apoyo directo en favor de la noción de verdad multivaluada, al menos un documento que señala la relación entre la noción comparativa de verdad y la teoría neoplatónica de la existencia.
En este trabajo nos proponemos mostrar, siguiendo algunas observaciones de H. Putnam, que la apelación al neoplatonismo no constituye un apoyo para la tesis de Háyek salvo que se acepte (de manera prejuiciosa) la superioridad de la lógica bivaluada.

Palabras claves: Predicado veritativo, Vaguedad, Semántica, Correspondentismo, Deflacionismo, Esquema (V)

1. INTRODUCCIÓN

En un trabajo reciente titulado "Ten questions and one problem on fuzzy logic" (preprint submitted to *Elsevier Science*, 22 July 1998), Petr Hájek relata su amistosa discusión con R. Parikh e intenta aislar ocho preguntas formuladas por Parikh para añadir otras dos y formular un comentario sobre cada una de ellas. La "Pregunta dos" es acerca de *cuáles son las propiedades importantes de los predicados vagos*. La respuesta de Hájek es que la propiedad más importante de tales predicados es su carácter comparativo: el hecho de que puedan aplicarse a un objeto (individual) dado en mayor o menor medida, incluyendo el sí absoluto (*absolutely yes*) y el no absoluto (*absolutely no*). Esta propiedad comparativa de los predicados vagos, sostiene, nos conduce a una "noción comparativa de verdad", en virtud de la cual una proposición puede ser "más verdadera" o "menos verdadera"; la misma proposición puede tener varios valores de verdad en diferentes situaciones: puede ser absolutamente verdadera, absolutamente falsa, o algo intermedio (*something in between*). Hájek observa que ha intentado preguntar a algunos filósofos si algo similar podría hallarse en la filosofía europea, pero la respuesta fue negativa. Finalmente, por casualidad, encontró un artículo de J. N. Martin ("Existence, Negation and Abstraction in the Neoplatonic Hierarchy", en *History and Philosophi of Logic*, 16, 1995, pp.169-196) acerca de la noción comparativa de existencia en los filósofos neoplatónicos, que Hájek considera, si no un apoyo directo en favor de la noción de verdad multivaluada, al menos un documento que señala la relación entre la noción comparativa de verdad y la teoría neoplatónica de la existencia. E inmediatamente destaca la importancia matemática de que la definición de la verdad de Tarski sea inmediatamente generalizable a las lógicas de predicados multivaluados. En efecto, dada una estructura L de valores de verdad, se toma un dominio nítido y se interpreta cada predicado P a través de una relación r_p L-valuada. La satisfacción de las fórmulas atómicas se generaliza inmediatamente: el valor de verdad de la fórmula P(x, ... , y) en una L-estructura dada y con una evaluación v de variables de objeto es r_p (v(x), ... , v(y)). Como algunos filósofos han señalado, dice Hájek, esta generalización inmediata se relaciona con lo que se suele llamar "esquema desentrecomillador" ("*dequotation scheme*"), según el cual "está nevando" es verdadera si yn sólo si está nevando, el valor de verdad de "John es alto" es el resultado de aplicar la función característica de altura al objeto llamado John.

2. DESARROLLO

Como puede advertirse, el texto de Hájek sugiere la existencia de un vínculo entre (i) la noción comparativa de verdad y la noción (neoplatónica) comparativa de existencia, y (ii) la noción de verdad y la definición tarskiana de verdad. De (i) se infiere que Hájek sostiene que (iii) hay un vínculo entre la noción de verdad y la noción de existencia, y de (ii) y (iii) se desprende que la tesis de Hájek al respecto es la siguiente: (iv) hay un vínculo entre la definición tarskiana de verdad y la noción de existencia. Esta argumentación de Hájek es particularmente interesante, ya que presenta una defensa, de la interpretación correspondentista de la así llamada "teoría semántica" de la verdad que Alfred Tarski expone en su artículo "The concept of Truth in Formalizad Languages" aparecido en versión inglesa en *Logic, Semantics, Metamathematics* en 1956.

Como señala E. Orlando (en "Sobre la interpretación deflacionaria de la teoría de Tarski", *Análisis Filosófico*, Buenos Aires, vol. XVII, 1997, N° 1, pp. 49-73), algunos autores sostienen que esta teoría debe interpretarse como una versión contemporánea de la teoría correspondentista tradicional que define la verdad como la correspondencia entre el lenguaje y el mundo: tal es el caso de Popper en *Conocimiento objetivo*, y de Donald Davidson en "True to the Facts" (*The Journal of Philosophy*, vol. LXVI, 1969, pp.748-764); mientras que para otros, en cambio, la teoría de Tarski representa una concepción deflacionaria según la cual la verdad se reduce a un expediente lingüístico para efectuar el ascenso semántico, es decir, el tránsito del uso de una oración que no contiene términos semánticos (por ejemplo, "la nieve es blanca") al uso de otra oración que sí los contiene (por ejemplo, "'la nieve es blanca' es una oración verdadera").

La concepción correspondentista y la deflacionaria tienen algo en común, en virtud del así llamado "criterio de adecuación material de una teoría de la verdad". Siguiendo a Tarski, diremos que satisfacer este criterio consiste en implicar, para cada oración de un lenguaje determinado, una oración como la siguiente: (1) "La nieve es blanca" es verdadera si y sólo si la nieve es blanca.

Esta oración ejemplifica el denominado "esquema (V)". Se trata de un bicondicional cuyo lado izquierdo atribuye el predicado veritativo a una oración del español (utilizando el nombre de esa oración, obtenido por el uso de las comillas) y cuyo lado derecho es esa misma oración, esto es, (1) ejemplifica la función desentrecomilladora del predicado veritativo para pasar de la mención de una oración al uso de la misma. Desde la perspectiva inversa, el predicado veritativo hace posible el ascenso semántico, esto es, el paso del uso de una oración que no contiene términos semánticos al uso de otra que sí los contiene. Como señala Orlando, el punto central consiste en que todas las oraciones como (1) parecen ser intuitiva y trivialmente verdaderas, ya que ningún hablante del español se negaría a aceptarlas, de manera que podemos pensar que el conjunto de todas ellas constituye el núcleo del concepto intuitivo de verdad.

Además, la concepción correspondentista se basa en otra idea que también se considera intuitiva, a saber, que la verdad involucra una cierta relación natural, de correspondencia o adecuación, entre el lenguaje y el mundo. La idea de correspondencia se entiende de la siguiente manera: ante todo, se trata de una relación asimétrica de fundamentación que va del mundo al lenguaje, esto es, el lenguaje se funda en el mundo y no a la inversa: la verdad exige que el lenguaje se adecue al mundo. Además, cada parte específica del lenguaje, es decir, cada oración particular, debe fundarse en una parte específica del mundo, o sea, en un hecho particular. Para utilizar el ejemplo anterior, la verdad de "la nieve es blanca" depende exclusivamente del hecho de que la nieve es blanca. Finalmente, el mundo con el cual se relaciona el lenguaje es objetivo e independiente de lo mental.

La concepción deflacionaria, por el contrario, se funda en la idea de que es la aceptación intuitiva de todas las instancias del esquema (V) y sólo eso lo que define el concepto de verdad. Así, el único fundamento para aceptar los bicondicionales (V) es el hecho de que sus lados izquierdos son cognoscitivamente equivalentes a sus lados derechos para cualquier hablante del castellano: afirmar los primeros es solamente una manera enfática de afirmar los segundos. De manera que la tesis central del deflacionismo es que el ascenso semántico (la función desentrecomilladora) no requiere para su cumplimiento de ninguna propiedad o relación sustantiva del tipo de la correspondencia. Así, el predicado veritativo no es considerado un predicado natural o físico sino un predicado lógico, como sostiene Paul Horwich en Truth (Oxford, Basil & Blackwell, 1990), o un operador de oraciones, como se sostiene en D. Grover- J. Camp y N. Belnap en "A Prosentential Theory of Thruth" (en *Philosophical Studies*, 27 (1.975), pp. 73-125).

De manera que, como puede advertirse, el punto (iv) de nuestra reconstrucción de la tesis de Hájek se debilita en la medida en que la teoría de la verdad de Tarski admite una interpretación alternativa a la correspondentista, esto es, la deflacionaria, por lo que el vínculo que se establece en (i), entre la noción comparativa de verdad y la noción comparativa de existencia, es innecesaria para fundamentar la teoría de los predicados vagos, como podría pensarse en primera instancia. Esta situación, entendemos, puede verse con mas claridad si acudimos a la presentación de la lógica trivaluada que Hilary Putnam ha expuesto en "Three-valued logic" (en *Mathematics, Matter and Method*, Cambridge University Press, Cambrige, 1985, pp. 166-173). Una lógica de este tipo acepta tres valores de verdad: verdadero, falso e "intermedio" ("middle"). El problema con una lógica de este tipo, observa Putnam, parece ser el siguiente: ¿tiene algún sentido la noción de un valor de verdad "intermedio"? Y suponiendo que el valor de verdad intermedio no signifique que una proposición tiene un valor de verdad desconocido (como suele interpretarse

generalmente el concepto de valor de verdad intermedio), pareciera que no podemos decir qué es lo que efectivamente significa (p.168). De manera de que muchos filósofos analíticos rechazan la posibilidad de una lógica trivaluada sobre la base de que no se puede ofrecer una traducción satisfactoria para la noción de tener un valor de verdad "intermedio". En efecto, sostiene Putnam, si la noción de ser una proposición con un valor de verdad intermedio se define explícitamente en términos de una lógica bivaluada, entonces lo que se obtiene es una interpretación trivial de la lógica trivaluada. Sin embargo, usar una lógica trivaluada -dice Putnam- tiene sentido en la medida en que usar una lógica trivaluada significa adoptar un modelo diferente de usar los términos lógicos. Hay un modo de usar los términos lógicos constitutivo de la lógica bivaluada ordinaria, y si estamos usando una lógica trivaluada nos comportaremos exactamente de la misma manera, excepto que emplearemos reglas propias de la lógica trivaluada y una definición trivaluada de "tautología". De manera que la expresión "usar una lógica trivaluada" significa adoptar un modo sistemático de utilizar los términos lógicos que sea coherente con el modo usual pero que entre en desacuerdo en ciertos casos (p.169).

3. CONCLUSIONES

Ahora bien, los filósofos analíticos están casi unánime de acuerdo en considerar a la lógica bivaluada como una lógica privilegiada, no solamente en el sentido de que correspondería al modo en que efectivamente usamos el lenguaje, sino especialmente porque no existen rivales desde un punto de vista lógico. Esto, sostiene Putnam, es un prejuicio del mismo tipo que el famoso prejuicio en favor del status privilegiado de la geometría euclideana, como cuando se sostiene que el espacio real tiene efectivamente tres dimensiones. Sin embargo, es posible pasar de un lógica bivaluada a una trivaluada sin cambiar completamente (*totlly*) el sentido de "verdadero" y "falso" y no en un sentido trivial como cuando se iguala "verdadero" con "alta probabilidad", "falso" con "baja probabilidad" e "intermedio" con "probabilidad intermedia". Este prejuicio ha llevado a sostener que el significado de "verdadero" ha sido suficientemente aclarado por Tarski para la lógica bivaluada, pero que no disponemos de una clarificación análoga para "verdadero" en los sistemas trivaluados. La respuesta obvia es que el famoso bicondicional "la nieve es blanca" es verdadero si y sólo si la nieve es blanca es perfectamente aceptable incluso si estamos usando una lógica trivaluada. En efecto, el criterio de Tarski tiene como consecuencia que se debe aceptar "la nieve es blanca " es verdadera si se acepta que la nieve es blanca, y que se debe rechazar "la nieve es blanca " es verdadera si se rechaza que la nieve es blanca, pero esto es solo el uso de "verdadero" y "falso" que puede perfectamente presentarse en una lógica trivaluada.

Si Putnam está en lo cierto, la apelación de Hájek a una antología neoplatónica solamente representaría un apoyo a favor de la teoría de los predicados vagos como predicados comparativos si se acepta el prejuicio en favor de la lógica bivaluada como lógica privilegiada.

Referencias

[1] Davidson, Donald: "True to the Facts", *The Journal of Philosophy*, vol. LXVI, pág. 748-764, 1969.

[2] Grover, D. - Camp, J. y Belnap, N.: "A Prosentential Theory of Thruth", *Philosophical Studies*, 27, pág. 73-125, 1975.

[3] Hájek, Petr: "Ten questions and one problem on fuzzy logic" preprint submitted to *Elsevier Science*, 22 July 1998.

[4] Harwich, Paul: *Truth*, Oxford, Basil & Blackwell, 1990.

[5] Martin, J. N. "Existence, Negation and Abstraction in the Neoplatonic Hierarchy", en *History and Philosophi of Logic*, 16, pág. 169-196, 1995.

[6] Orlando, E.: "Sobre la interpretación deflacionaria de la teoría de Tarski", *Análisis Filosófico*, Buenos Aires, vol. XVII, Nº 1, pág. 49-73, 1997.

[7] Popper, Karl: "Conocimiento objetivo", *The Journal of Philosophy*, vol. LXVI, pág. 748-764, 1969.

[8] Putnam, Hilary: "Three-valued logic", *Mathematics, Matter and Method*, Cambridge University Press, Cambrige, pág. 166-173, 1985.

[9] Tarski, Alfred: "The concept of Truth in Formalizad Languages", *Logic, Semantics, Metamathematics*, 1956.

Fuzzy logic abduction

Peter Vojtáš[1]

Abstract. Abduction is an important form of nonmonotonic reasoning allowing one to find explanations for certain symptoms or observations. We consider the situation when the application domain is described by a logic program (IF-THEN rules). To model uncertainty in human cognition and real world applications, we use fuzzy logic (i.e. many valued logic where axioms can have truth values (confidence factors) smaller than 1). Our connectives are tunable to fit real world data. We introduce a model of fuzzy logic programming abduction problem FLPAP which has many applications in diagnostical systems. We show soundness and completeness of declarative and procedural semantics of FLPAP. We show how can linear programming be applied in order to get minimal solutions wrt a cost function. We discuss the problem of approximative explanations in finitely valued logic and integer programming.

1 Introduction

This paper presents a model of abduction under uncertainty based on fuzzy logic programming. Abduction is an important form of nonmonotonic reasoning allowing one to find explanations and/or causes for certain symptoms or observations (introduced by C. S. Peirce in [8]).

We restrict ourselves to logic based abduction, especially to abduction applied to logic programming (see [3], [6]). To model uncertainty in human cognition and real world applications we offer the possibility to use fuzzy logic (see [5]).

2 Example

2.1 The application domain

The following illustrative example from the domain of motor vehicles is inspired by [2] and [3] (subscripts L, P, G denote connectives of Łukasiewicz, product and Gödel logic)). Consider the following **fuzzy logic program** P_{mv} (fuzzy IF-THEN rule base) in a prolog syntax with attached truth values (which can be also understood as a fuzzy set of rules, i.e. a partial mapping P_{mv} : Rules $\longrightarrow (0,1]$)

high_fuel_consumption $\longleftarrow_G$ rich_mixture$\wedge_L$ low_oil; 0.8.
overheating $\longleftarrow_P$ low_oil; 0.5.
noisy_behaviour $\longleftarrow_P$ rich_mixture; 0.8.
overheating $\longleftarrow_L$ low_water; 0.9.
noisy_behaviour $\longleftarrow_P$ low_oil; 1.

Denote the set of all propositional variables describing our universe of discourse Var_{mv}.

2.2 Observations

If we observe that our car is noisy, overheated and has a high fuel consumption, we would like to know why it is so. Let us call these variables **observation variables** and denote it OV_{mv} = { noisy_behaviour, overheated, high_fuel_consumption}. The second parameter of our abduction problem are **observations** (sometimes called manifestations, symptoms, effects) represented by a fuzzy theory $OBS_{mv} : OV_{mv} \longrightarrow [0,1]$ consisting of facts=observation variables equipped with truth values

high_fuel_consumption; 0.25
overheating; 0.25.
noisy_behaviour; 0.5

2.3 Hypothesis and explanations

Possible explanations will be fuzzy subsets of the set of **hypothesis** H_{mv} = { rich_mixture, low_oil, low_water }. This makes sense, because in the realm of a theory and of observations suffering from uncertainty, we can expect that certain level of confidence of hypothesis can be (under the presence of the theory) an explanation of these (uncertain) observations.

A fuzzy set $E : H_{mv} \longrightarrow [0,1]$ is an **explanation** of our fuzzy logic programming abduction problem (FLPAP) $\mathcal{A}_{mv} =< P_{mv}, OBS_{mv}, H_{mv} >$ given by the theory P_{mv}, observations OBS_{mv} and the set of hypothesis H_{mv}, if the theory $P_{mv} \cup E$ is consistent and $P_{mv} \cup E$ semantically implies all observations OBS_{mv}. In our case FLPAP $\mathcal{A}_{mv}$ has an infinite set of solutions, it contains especially the explanation which takes values constantly 1 (we do not have negation in our example). The fuzzy theory E_2 defined as follows E_2(rich_mixture) = E_2(low_oil) = 0.7 and E_2(low_water) = 0 is also an explanation of FLPAP $\mathcal{A}_{mv}$. Indeed, for an arbitrary truth assignment of all propositional variables $I : Var_{mv} \longrightarrow [0,1]$ holds: if I is a model of both P_{mv} and E_2 then we can calculate that for all $m \in OV_{mv}$ is $I(m) \geq OBS_{mv}(m)$. Because I is a model of E_2 we have e.g. I(low_oil) $\geq E_2$(low_oil) = 0.7 (using truth value functions for connectives) we get from the second rule that either

[1] Institute of Computer Science, Academy of Sciences of Czech republic, Pod vodárenskou věží 2, 18207 Prague, Czech republic, email: vojtas at uivt.cas.cz, Dpt. Comp. Sci. University of PJŠafárik Košice, Slovakia, partialy supported by COST 15 and VEGA 1/4375/97

$I($ overheating $) \geq I($low_oil$) \geq 0.7$ or

$$0.5 \leq \frac{I(\text{ overheating })}{I(\text{ low_oil })} \leq \frac{I(\text{ overheating })}{0.7}$$

hence in both cases is $I($ overheating $) \geq 0.35 > 0.25 = OBS_{mv}($ overheating $)$ (for other observations similarly). □

3 Fuzzy logic programming

3.1 The language

Our **language** of propositional fuzzy logic consists of a finite set Var of propositional (sentential) variables and several rationality preserving and definable connectives: conjunctions $\&_1, \&_2, \ldots, \&_k$, disjunctions $\vee_1, \vee_2, \ldots, \vee_l$, implications $\rightarrow_1, \rightarrow_2, \ldots, \rightarrow_m$ and the negation $\neg$. For a connective c the corresponding semantical counterpart - the truth value function will be denoted by c^* (i.e. with a dot over the very connective, e.g. $\neg^*(x) = 1 - x$). Conjunctions are assumed to be t-norms, disjunctions are t-conorms ([7]). We assume that for all implications in our language there is a generating t-norm T such that $\rightarrow_T^* (u, v) = \sup\{w : T(u, w) \leq v\}$ (see e.g. [4], [5]). Note that T need not be a truth value of any connective in our language (e.g. in our example $\&_P^*$ is generating $\rightarrow_P$, although $\&_P$ does not appear in our program).

3.2 Fuzzy theories

A **fuzzy theory** is a fuzzy set of propositions Th : Formulas $\longrightarrow [0, 1] \cap Q$ assigning each proposition a rational number. A single axiom will be denoted as a pair $(C; Th(C))$ (see [5]). **Semantical interpretations** of our language $I : Var \longrightarrow [0, 1]$ are assignments of variables with truth values (real numbers, [5]). A truth value assignment I is a **model of a fuzzy theory** Th if for all formulas C is $Th(C) \leq I(C)$. This definition says, that a theory and data fit together if the expert given truth value is a lower bound for truth values from data. Of course it is interesting to have the strongest theory fitting data (completeness problem). Another problem is whether our data are truth value functional (if not we have to use other approaches e.g. probabilistic, belief functions,..., see e.g. [1]).

3.3 Fuzzy logic programs

In general, a formula B is called a **body** if it is in disjunctive normal form. A **rule** is a proposition of form $A \longleftarrow B$, where A is a propositional variable and B is a body. A propositional variable is also called a **fact**. A **fuzzy logic program (FLP)** is a fuzzy theory P : Rules $\cup$ Facts $\longrightarrow (0, 1] \cap Q$ such that the domain of P is finite and consists of rules and facts. A propositional variable can be intended as a **query** (in this case it is denoted by $A?$ (note, we do not adopt refutation here). For a fuzzy logic program P and a query $A?$ a **correct answer** is a real number $x \in [0, 1]$ such that for all truth assignments (semantic interpretations) I the following holds: If I is a model of the fuzzy theory P then $I(A) \geq x$.

3.4 Procedural semantics of FLP

As far as our connectives are subject for tuning and/or learning we are in a logical theory in which we (possibly) do not have any logical axioms. Motivation for our procedural semantics is based on the "backwards" understanding of many-valued modus ponens (do not confuse it with so called fuzzy modus ponens which is an interpolation and/or extrapolation deduction rule for deduction by analogy). Many valued modus ponens is a sound rule ([4], [5], [9]) which looks as follows:

From	$(B \longrightarrow_T A; y)$
and	$(B; x)$
infer	$(A; T(y, x))$.

In this case we say that the generating t-norm T evaluates modus ponens. The backward understanding argues as follows: Querying A, we look for an implication with A in the head, choose one of them, say $(B \longrightarrow_T A, y)$. From now on we know that the truth value of our answer will have value $T(y, b)$ for some b, where b will be one of (in our case (B, x)) computed truth values of B. We can write our backward modus ponens inference as

$A?$	deduce by $(A \longleftarrow_T B; y)$ to
$T(y, B)$	deduce by $(B; x)$ to
$T(y, x)$	

3.5 Admissible rules

Our inference rules are in the form of a context free grammar, where X and Y are variables for words (possibly empty).

Rule 1. From XAY infer $XT(P(A \leftarrow_T B), B)Y$

Rule 2. From XAY infer $XP(A)Y$

Rule 3. From $XB_1 \& B_2Y$ infer $X\&^*(B_1, B_2)Y$

Rule4a. From $XB_1 \vee B_2Y$ infer $X \vee^* (B_1, B_2)Y$

Rule4b. From $XB_1 \vee B_2Y$ infer $X \vee^* (B_1, 0)Y$

Rule4c. From $XB_1 \vee B_2Y$ infer $X \vee^* (0, B_2)Y$ (we need these because of possibility one member of disjunction fails and then there is no aggregation of confidences of single diagnosis)

Rule 5. From $X \neg BY$ infer $X1 - \max(b_1, \ldots, b_n)Y$ if the computational tree for P and $B?$ is finite and $b_1, \ldots, b_n$ are calculated values of all successful branches (failed branches are interpreted to have value 0).

Rule 6. If there are in the word X no letters denoting propositional variables (the inference was successful), then the word X is a composition of truth value functions for connectives and evaluating t-norms with rational arguments, then from X infer the value of that expression.

For a fuzzy program P and a query $A?$ we say that a rational number $q \in [0, 1]$ is a **computed answer** for P and $A?$ if there is a sequence $G_0, G_1, \ldots, G_n$ (successful computation) such that $G_0 = A$, $G_n = q$ and for all $i < n$, G_{i+1} is inferred from G_i by one of admissible rules.

In [9] there is developed a fixpoint semantics for definite fuzzy logic programs (i.e. without negation and hence without Rule 5) and soundness and approximative completeness of procedural semantics is proved (for lower semicontinuous conjunctions, disjunctions and generating t-norms).

Rule 5 gives consistent results for our fuzzy logic programs, but it is out of the scope of this paper to discuss a semantics for it.

4 Fuzzy abduction

4.1 Definition of the problem

For two fuzzy theories T and S denote by $T \cup S$ the fuzzy theory defined by $(T \cup S)(A) = \max\{T(A), S(A)\}$.

Definition 1 *A* **fuzzy logic programming abduction problem (FLPAP)** *consists of a tuple $\mathcal{A} =< P, OBS, H >$, where $H \subseteq Var$ is the set of hypothesis, $OBS : OV \longrightarrow [0,1]$ is the fuzzy theory of observations (where OV is the set of observation variables such that $OV \cap H = \emptyset$) and P is a fuzzy logic program (where intended meaning is that observation variables should not be explained by themselves).*

Definition 2 *A fuzzy theory $E : H \longrightarrow [0,1]$ is a* **correct explanation** *to FLPAP $\mathcal{A} =< P, OBS, H >$ if*

(1) $P \cup E$ semantically implies OBS, i.e. for every semantical interpretation (truth assignment of variables) $I : Var \longrightarrow [0,1]$ the following holds: If I is a model of $P \cup E$ then I is a model of OBS.

(2) $P \cup E$ is consistent

It will be useful to represent solutions as a subset of many dimensional unit cube. Namely, if $H = \{h_1, \ldots, h_n\}$ is the set of hypothesis and $E : H \longrightarrow [0,1]$ is a solution, it is uniquely determined by its values $(E(h_1), \ldots, E(h_n)) \in [0,1]^n$. So having an element of nth power of unit cube $e = (e_1, \ldots, e_n) \in [0,1]^n$ it represents the mapping $E_e(h_i) = e_i$. Denote by $\mathbb{SOL}_d(\mathcal{A})$ the set $\{e \in [0,1] : E_e$ is a correct explanation for FLPAP $\mathcal{A}\}$ (the subscript d remembers us the definition is a declarative one)

4.2 Existence of explanations

Eiter and Gottlob have shown in [3] that deciding $\mathbb{SOL}_d(\mathcal{A}) \neq \emptyset$ is complete for complexity classes at the second level of polynomial hierarchy, while the use of priorisation raises the complexity to the third level in certain cases (for arbitrary propositional theories). We show that for definite FLPAP $\mathcal{A}$ are both the existence and priorisation (minimality) in the first level (NP).

Continuation of example. Having our motor vehicle example from 2.1. and our fuzzy program P_{mv} and the query high_fuel_consumption? by fuzzy logic program computation using first program rule we get

$$\min(0.8, \max(0, \text{rich_mixture} + \text{low_oil} - 1))$$

Now similarly as in the two valued logic abductive logic programming ([6]) our procedure instead of failing in a proof when a selected subgoal fail to unify with the head of any rule, the subgoal is viewed as a hypothesis, that is, if we know truth value for low_mixture and low_oil (from an explanation) we have the computed answer for high_fuel_consumption. To fulfill OBS_{mv}(high_fuel_consumption) ≥ 0.25 it should be $\min(0.8, \max(0, \text{rich_mixture} + \text{low_oil} - 1)) \geq 0.25$ hence rich_mixture + low_oil ≥ 1.25 □

So our fuzzy logic programming abduction should run as a usual logic program with two exceptions

- it successfully ends without deducing variables which are in the set of hypothesis
- it is prompted by a query with threshold, which can serve as a cut (as we will see later).

Definition 3 *(Procedural semantics for FLPAP) Let $\mathcal{A} =< P, OBS, H >$ be a FLPAP and let $\mathcal{G} = (G_0, G_1, \ldots, G_l)$ be a sequence of words in the alphabet of fuzzy logic program computation. We say that the sequence $\mathcal{G}$ is a* **successful abduction** *for $\mathcal{A}$ and $m \in OV$ if*

$G_0 = m$,

G_l contains only variables from H and

for all $i < l$, G_{i+1} is inferred from G_i by one of admissible rules and for the constantly one interpretation $I_1 : Var \longrightarrow \{1\}$ is the truth value $I_1(G_{i+1}) \geq OBS(m)$ (this condition is to be understood as a cut, because it estimates the best possible computation)

For $H = \{h_1, \ldots, h_n\}$, the expression G_l can be understood as a function of n real variables $G_l : [0,1]^n \longrightarrow [0,1]$ and that is why we can denote it by $G_l = f_{\mathcal{G}}(h_1, \ldots, h_n)$.

Theorem 4 *(Existence od solutions) Let $\mathcal{A} =< P, OBS, H >$ be a FLPAP and for each $m \in OV$ there is a successful abduction for $\mathcal{A}$ and m. Then $\mathbb{SOL}_d(\mathcal{A}) \neq \emptyset$.*

Proof: Because of our cut, which in each step checks whether substituting truth value 1 still keeps $I(G^m_{i+1}) \geq OBS(m)$, especially for the last step, hence for all $m \in OV$ is $f_{\mathcal{G}^m}(1, \ldots, 1) \geq OBS(m)$ □

4.3 Correct explanations versus computed explanations

Our definition of abduction gives us the possibility to define computed explanations for FLPAP $\mathcal{A}$.

Definition 5 *A tuple $e = (e_1, \ldots, e_n)$ is a* **computed explanation** *for a FLPAP $\mathcal{A} =< P, OBS, H >$ if for every $m \in OV$ there is a successful abduction $\mathcal{G}^m$ for $\mathcal{A}$ and m such that $f_{\mathcal{G}^m}(e_1, \ldots, e_n) \geq OBS(m)$ The set of all computed explanations will be denoted by $\mathbb{SOL}_p(\mathcal{A})$.*

Subscript p resembles us to procedural character of this definition

Example continued. In our motor vehicle example we calculated in 4.2 that first rule of P_{mv} gives rich_mixture + low_oil ≥ 1.25, similarly second rule gives low_oil ≥ 0.5 and the third gives rich_mixture ≥ 0.625, hence the set (coordinates are ordered as (rich_mixture, low_oil, low_water))

$$\{(e_1, e_2, e_3) \in [0,1] : \quad e_1 + e_2 \geq 1.25 \text{ and } e_1 \geq 0.625 \text{ and } e_2 \geq 0.5\}$$

is a subset of $\mathbb{SOL}_p(\mathcal{A})$ □

Theorem 6 *(Soundness of fuzzy abduction) Assume $\mathcal{A}$ is a definite FLPAP, then $\mathbb{SOL}_p(\mathcal{A}) \subseteq \mathbb{SOL}_d(\mathcal{A})$ (that is, every computed explanation for $\mathcal{A}$ is also a correct explanation)*

Proof: Our inverse usage of modus ponens is sound and so is our threshold cut (because of monotonicity of t-norms

and conorms). So the result follows by induction through the length of abduction. □

In the next completeness theorem we need the assumption that our logical program has a finite computational tree - according to abductions. This is in practical applications of abduction very often the case, because e.g. observations should not be explained by themselves (i.e. there is no recursion on rules defining them), and most of logic programs for abduction are layered. Moreover if t-norms are archimedian (also very often) then the iteration of t-norms ends below the observation value threshold, and hence is cut.

Theorem 7 *(Completeness of fuzzy abduction) Assume $\mathcal{A}$ is a definite FLPAP and the logical program has a finite computational tree according to abductions, then $\mathbb{SCL}_d(\mathcal{A}) \subseteq \mathbb{SCL}_p(\mathcal{A})$ (that is, every correct explanation for $\mathcal{A}$ is also a computed explanation)*

Proof: Let $e = (e_1, \ldots, e_n)$ be a correct explanation for FLPAP $\mathcal{A} =< P, OBS, H >$. Recall that it uniquely determines an explanation $E_e : H \longrightarrow [0,1]$. Consider the fuzzy logic program P_e which consists of rules of the program P and of facts $h_i; e_i.$ for all i. The fact that e is a correct explanation for $\mathcal{A}$ means exactly that for all $m \in OV$ is $OBS(m)$ a correct answer for the program P_e and query $m?$. Hence (by approximative completeness from [9]) for every $\epsilon > 0$ there is a successful prolog computation $\mathcal{G}_\epsilon^m$ such that the computed answer

$$q_\epsilon^m = f_{\mathcal{G}_\epsilon^m}(e_1, \ldots, e_n) \geq OBS(m) - \epsilon.$$

Choose an infinite sequence $\epsilon_r \longrightarrow 0^+$. As far as there are only finitely many computations, there is between all $\mathcal{G}_{\epsilon_r}^m$'s a single computation $\mathcal{G}_\delta^m$ which does the job for all but finitely many ϵ_r's, hence

$$q_\delta = f_{\mathcal{G}_\delta^m}(e_1, \ldots, e_n) \geq OBS(m) - \epsilon_r.$$

for all but finitely many r's hence $q_\delta \geq OBS(m)$. Hence if $OV = \{m_1, \ldots, m_k\}$ then for every $i < k$ there is a computation $\mathcal{G}^k$ such that

$$q_k = f_{\mathcal{G}^k}(e_1, \ldots, e_n) \geq OBS(m),$$

so for every k is $\mathcal{G}^k$ a witness for q_k being a computed answer for P_e and m_k. Hence e is a computed explanation. □

From now on we do not have to distinguish between two sets of solutions and we simply denote it $\mathbb{SCL}(\mathcal{A})$.

One can ask how difficult is to decide whether $e \in \mathbb{SCL}(\mathcal{A})$ or not. Now we see that it substantially depends on the complexity of logic programming computation and the complexity of functions evaluating the truth values for connectives and evaluating t-norms. So from a computational point of view, it makes sense to use simple connectives (e.g. linear in each coordinate (as product is) or even partly constant).

4.4 Cheapest explanation via linear programming

In real world applications one can be interested in cheapest/minimal solutions of the abduction problem.

Example continued. Continuing our example on motor vehicles assume, that to check (and fix) an explanation $(e_1, e_2, e_3) \in \mathbb{SCL}(\mathcal{A}_{mv})$ costs $2e_1 + e_2 + 0.1e_3$. The space of solutions $\mathbb{SCL}(\mathcal{A}_{mv})$ is bounded by linear surfaces in $[0,1]^3$ and is union of 4 convex bodies.

One of them we get using first three rules of the program P_{mv}. Applying a linear programming method for minimal solution to the set

$$\{(e_1, e_2, e_3) \in [0,1] : \quad e_1 + e_2 \geq 1.25 \text{ and } e_1 \geq 0.625 \text{ and } e_2 \geq 0.5\}$$

wrt our cost function we get in this convex body a minimal solution solution $(0.625, 0.625, 0)$ at cost of 1.875.

The cheapest solution of FLPAP $\mathcal{A}_{mv}$ is $e_{min} = (0.25, 1, 0.35)$ at cost of 1.535 (we get it from the convex body of all solutions to first, fourth and fifth program rule). □

As linear programming is lying in NP complexity class (even much lower) as prolog does, to find minimal solutions for a definite FLPAP (assuming connectives and t-norms are coordinatewise linear) does not increase the complexity and remains in NP.

4.5 Finite approximations

In many real world applications we work with approximations of data and measurements are interesting up to some precision. We can show that our results work also for connectives which are partly constant lower semicontinuous functions (we loose associativity of connectives), our fuzzy logic is finitely valued logic, fuzzy abduction is easy to implement and finding cheapest (minimal) solutions restricts to a problem of integer programming.

REFERENCES

[1] D. Dubois, J. Lang, H. Prade. Fuzzy sets in approximative reasoning, part2, Logical approaches. Fuzzy Sets and Systems 40 (1991) 203–244

[2] L. Console, D. Thesier Dupre, P. Torasso. On the relationship between abduction and deduction. J. Logic Comput. 1 (1991) 661–690

[3] T. Eiter, G. Gottlob. The complexity of logic based abduction. J. ACM 42(1995)3–42

[4] S. Gottwald. Fuzzy sets and fuzzy logic. Vieweg, Berlin, 1993

[5] P. Hájek. Metamathematics of fuzzy logic. Kluwer, Dodrecht, 1998

[6] A. C. Kakas, R. A. Kowalski, F. Toni. Abductive logic programming. J. Logic Computat. 2(1992)719–770

[7] R. Mesiar. Fuzzy sets and probability theory. Tatra Mt. Math. Publ. 1 (1992) 105–129

[8] C. S. Peirce Abduction and induction. In Philosophical writings of Peirce, chapter 11, J. Buchler ed. Dover, New York 1955

[9] P. Vojtáš. Fuzzy reasoning with tunable t-operators. J. Adv. Comp. Intelligence 2(1998)121–127

THE DISTRIBUTION PROBLEM IN MANAGEMENT ACCOUNTING WITH GENETIC ALGORITHM AND FUZZY SETS

Enrique López González, Miguel A. Rodríguez Fernández, Cristina Mendaña Cuervo and Raquel Flórez López
Department of Economy and Business Management
University of León
24071-León, Spain
dde{elg/mrf/cmc/rfl}@unileon.es

1. ABSTRACT

Many different optimisation problems can be formulated on networks: For instance, we might be interested in finding the shortest path from one node to another, finding the cheapest way to connect all the nodes, or finding the best way to move objects through a network. The movement of objects through a network is called a flow.

In this paper we pretend to show a new approach to a classical network optimisation problem: the transportation problem. Firstly, allowing the vehicles that can establish routes, and, also, permitting to manage with imprecise or vague information. We propose to use the Fuzzy Sets Theory [9]; [4]; [10], with the aim of being able to handle the uncertainty, which is characteristic of decision-making processes in distribution problems.

Moreover, to optimise the distribution network we suggest to use a Genetic Algorithm (GA) [3]; [1]; [7]. The main reason for this is that GAs are heuristic optimisation methods which don't impose restrictions to the posing of the problem. In this study, the algorithm is characterised by the use of a Fuzzy Fitness Function that allows the evaluation of imprecise information.

Keywords: transportation and distribution problems, shortest route problem, fuzzy sets, genetic algorithms.

2. FUZZY DISTRIBUTION MODEL

The information necessary is related to:

- *Destination or client demands ($\tilde{C}$).* We consider a more realist representation using fuzzy sets.

- *Sources number and its supply capacity (F).* We suppose that each source has a supply capacity, normally known in a strict way.

- *Vehicle Capacity (V).* We consider that in each source the company has a vehicle used to distribute the commodities. Therefore, we must take into account the vehicle capacity, because the number of travels will depend of it.

- *Transporting costs of the empty vehicles (cv)* Sometimes, the vehicles must return to their sources due to they have distribute all their capacities. Thus, we must take into account what is the shipping cost when the vehicle is empty. This, normally depends of the distance (routed kilometres).

- *Increasing cost for each unit of commodities shipped ($\Delta c\tilde{v}$).* When the vehicles distribute the commodities the transporting cost increase with each unit shipped. As well, this increasing cost depends of the distance travelled. Moreover, this cost could be represented using fuzzy sets.

- *Distance among sources and destinations (D).* The distribution is made from the sources to the destinations or clients. So, it is necessary to know the distance among them.

- *Distance among destinations (D').* The vehicles can establish routes from the source to the several destinations that this source supplies, so it is necessary to know the distances among the destinations.

According with this approach, we are trying to reach the real distribution problem. Thereafter, we need some tool capable to solve this complex problem or able to give a reasonable solution. In this paper we propose to use a Genetic Algorithm with a Fuzzy Fitness Function. Also, although we describe to use fuzzy sets, we suggest representing all the fuzzy information as Trapezoidal Fuzzy Numbers (TFN) [5].

3. A GENETIC ALGORITHM FOR FUZZY DISTRIBUTION PROBLEMS

Some authors have applied Genetic Algorithms to transportation problems, in many cases solving the problem reducing their complexity or using precise knowledge for the variables.

In our approach we suggest a fuzzy representation of the information on held and a less restrictive model. Due to this, the GA must determine the quantities supplied from each source to each destinations and the routes travelled for the vehicles. For both objectives the criteria is the same: minimising the total distribution costs.

To solve the fuzzy distribution problem we propose to use a GA with the following components:

3.1. Genetic Representation

The solutions to the problems are a set of distribution quantities to be shipped from each source to each destination and the routes followed by the vehicles of each source. To codify this solutions we propose to use two matrixes: one that contains the quantities distributed from each origin to each destination, and, other, that represents the path followed for the several vehicles.

The codification of quantities matrix, for an example of five sources (S_i) that must supply to four destinations (D_i) could be:

Table 1: Codification of quantities matrix

S_1^1	S1	S2	S3	S4	S5	**Total**
D1	15	0	0	70	0	**85**
D2	0	40	0	0	0	**40**
D3	25	10	40	0	50	**125**
D4	0	0	60	0	5	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

that indicates the quantities supplied from each source to each destination.

On the other hand, the codification of vehicles routed could be:

Table 2: Codification of vehicles matrix

S_1^2	S1	S2	S3	S4	S5
D1	1	2	1	4	3
D2	4	3	2	3	2
D3	2	1	3	2	1
D4	3	4	4	1	4

that indicates the position of each destination in the vehicle routes.

Moreover, in order to generate usefulness solutions, the company must decide the covering level of each client fuzzy demand. The TFN can be identified as possibility distributions. So, we establish in the work the risk of not cover demand of each client. According to this, the quantity matrix represents available solutions for the decision.

3.2. Fuzzy Fitness Function

The fitness function must assign more value to good solutions to the distribution problem. To get this, we propose to calculate the total distribution cost that the two matrixes suppose. The steps are as follows:

First, each vehicle must depart from its source full of commodities. The first destination established for the routes matrix determine the cost of this first travel. Once the first destination is satisfied (with one or more travels) the second destination is attempted. If the vehicle is empty during the route it return empty to the source to full its capacity. As well, when all the destinations are satisfied then the vehicle go back to its origin, completing the route. The sum of all these transporting costs would be a part of the solution. To obtain the total cost we must add the cost of the other distribution routes contained in the matrixes.

Due to the fact that some of the information can be represented by fuzzy sets, we suggest to use the operations designed for them [2]. Moreover, in the cases that the operations could not produce a TFN, we have approximated the result as one of this [5], assuming a little error.

To set up a hierarchy among the solutions, the proposal is to use the fuzzy distance [6], based on Hamming Distance. Doing it, we calculate the distance from the origin (singleton 0) to each solution fuzzy cost, which is defined as follows:

$$d(\widetilde{A},\widetilde{B}) = \int_{\alpha=0}^{1} \left(\left| A_{\alpha}^{1} - B_{\alpha}^{1} \right| + \left| A_{\alpha}^{2} - B_{\alpha}^{2} \right| \right) d\alpha \qquad (1)$$

where $[A_{\alpha}^{1}, A_{\alpha}^{2}]$ is confidence interval of $\widetilde{A}$ at the signification level α.

The more accurate solutions will have a lower distance. Thus for obtaining the fitness value we propose calculate the inverse of the distance of each solution.

3.2.1. Selection process

We propose to use Roulette Wheel Ranking [3] to select the "parents" of the next generation.

3.2.2. Crossover operator

Traditional crossovers cannot be used, because the solutions are two matrixes that have to accomplish some

constraints. Due to this we propose to different crossovers for each matrix, as follows:

♦ **Distribution quantities matrixes.** To combine the information of these matrixes obtained from two "parents" we have used the crossover proposed for Vignaux and Michalewicz to Transportation Problem [8]:

• Continuing with the aforementioned example of five sources that must supply to four destinations. Lets imagine that after selection process we have two quantities matrixes to be crossed like these:

Tables 3 and 4: Matrixes to be crossed

S_1^1	S1	S2	S3	S4	S5	**Total**
D1	15	0	0	70	0	**85**
D2	0	40	0	0	0	**40**
D3	25	10	40	0	50	**125**
D4	0	0	60	0	5	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

S_2^1	S1	S2	S3	S4	S5	**Total**
D1	0	0	50	0	35	**85**
D2	0	40	0	0	0	**40**
D3	5	0	50	70	0	**125**
D4	35	10	0	0	20	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

• Using these "parents", we obtain two temporal matrixes: one containing the integer average of both "parents" (*MED*), and another that include an 1 in those cells that have been rounded to obtain the average, and a 0 in the remaining (*RED*).

• The *RED* matrix have some interesting properties. One of then indicates that the number of *ones* of each row or column is a par one. In other words, the source capacity values and the destination demand values are par integers. Thus, we can divide the *RED* matrix into two matrixes (*RED1* and *RED2*):

$$RED = RED1 + RED2, \qquad (2)$$

being equal the supplying capacity of *RED1* and *RED2* for each source, and, also, the demands of *RED1* y *RED2* for each destination. The offspring are obtained as follows:

$$S'^1 = MED + RED1$$
$$S'^2 = MED + RED2 \qquad (3)$$

Then, for the example, the "sons" could be:

Tables 5 and 6: "Sons"

$S_1^{1'}$	S1	S2	S3	S4	S5	**Total**
D1	8	0	25	35	17	**85**
D2	0	40	0	0	0	**40**
D3	15	5	45	35	25	**125**
D4	17	5	30	0	13	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

$S_2^{1'}$	S1	S2	S3	S4	S5	**Total**
D1	7	0	25	35	18	**85**
D2	0	40	0	0	0	**40**
D3	15	5	45	35	25	**125**
D4	18	5	30	0	12	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

♦ **Vehicle routes matrixes.** As well, we must cross the route matrixes of both "parents". To this, we first select randomly one source and interchange between the "parents" the routes that the vehicle of this origin must follow.

3.3. Mutation operator

The intention of this operator is to introduce diversity into the solutions. Trying to this we must make differences among the two types of matrixes, as follows:

♦ **Distribution quantities mutation.** We propose to use a special mutation method designed to keep the solutions resulting as feasible ones. The steps are as follow:

• *Step 1.* We select one source (Source A) and one destination (Destination A) of the matrix. This destination must receive a non-zero quantity from the source selected. For the example, if we mutate the first matrix obtained after the crossing process, the source and destination could be:

Table 7: Selection of source A and destination A

S_1^1	S1	S2	S3	**S4**	S5	**Total**
D1	8	0	25	35	17	**85**
D2	0	40	0	0	0	**40**
D3	15	5	45	**35**	25	**125**
D4	17	5	30	0	13	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

• *Step 2.* We generate a random number between 0 and the quantity assigned to this destination. For the example could be 20.

• *Step 3*. We search in the remaining sources, one quantity assigned to a different destination which amount is bigger that the random number generated before (Source B and Destination B). For the example could be:

Table 8: Selection of source B and destination B

S_1^1	S1	**S2**	S3	S4	S5	**Total**
D1	8	0	25	35	17	**85**
D2	0	**40**	0	0	0	**40**
D3	15	5	45	35	25	**125**
D4	17	5	30	0	13	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

• *Step 4*. We discount the random number to the assigned amounts from the Source A to Destination A and from the Source B to Destination B. After that, we add the random number to the assigned amounts from Source A to Destination B and from Source B to Destination A. According to this, for the example, the mutated matrix could be:

Table 9: The mutated matrix

S_1^{11}	S1	**S2**	S3	**S4**	S5	**Total**
D1	8	0	25	35	17	**85**
D2	0	**20**	**20**	0	0	**40**
D3	15	**25**	**25**	35	25	**125**
D4	17	5	30	0	13	**65**
Total	**40**	**50**	**100**	**70**	**55**	**315**

Finally, with the aforementioned process we can mutate the quantity matrixes maintaining them as feasible ones.

♦ **Vehicles routes mutation**. On the other hand, for mutating the route matrix we propose an *interchange* method. First, we select randomly one source. After that, we select two destinations and interchange their position in the route followed for the source vehicle.

3.4. Halt criteria for the best solution search

The proposal is for the algorithm to go through a number of generations specified by the user until the best solution is found. Moreover, in order not to lose good solutions, the characteristic termed *elitism* [3] has been introduced. This procedure consists of keeping the best individual from a population in successive generations unless and until some other individual succeeds in doing better in respect of suitability. In this way, the best solution is not lost until outclassed by a more suitable solution.

As explained, application of the model proposed here allows solving the distribution problem under uncertain environmental conditions.

4. CONCLUDING REMARKS

There are two main results obtained from this paper. One is the formulation of an approach to Distribution Problem that could be more adapted to real situations. The other is showing the GA capacity to solve complex problems with imprecise or vague information. In that way the work group has developed an example that deals with the distribution problem for a production furniture company. The solution has been found using the GA proposed.

However, trying to adjust properly the posing of the problem, as proposal of future work, could be interesting take into account the accomplishing of production and demand cadences, multiple products distribution, the intermediate warehouses, etc.

References

[1] V. Biethahn and V. Nissen, *Evolutionary Algorithms in Management Application*, Springer-Verlag (1995).

[2] D.Dubois and H. Prade. *Fuzzy Sets and Systems: Theory and Applications*. Academic Press, New York (1980).

[3] D.E. Goldberg. *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison Wesley (1989).

[4] A. Kaufmann. *Introduction to the Theory of Fuzzy Subsets*. Academics Press, New York (1975).

[5] A. Kaufmann and J. Gil Aluja. *Introducción de la teoría de los subconjuntos borrosos en la Gestión de Empresas*. Milladoiro, Santiago de Compostela (1986).

[6] A. Kaufmann and J. Gil-Aluja. *Técnicas Operativos de Gestión para el Tratamiento de la Incertidumbre*. Hispano Europea, Barcelona (1987).

[7] E. López-González, C. Mendaña-Cuervo and M.A. Rodríguez Fernández, The Election of a Portfolio Through a Fuzzy Genetic Algorithm: The Pofugena Model, *Operational Tools in the Management of Financial Risks*, Kluwer Academic Publishers, pp. 273-290 (1998).

[8] G.A. Vignaux and Z. Michalewicz. A Genetic Algorithm for the Linear Transportation Problem. IEEE Transactions on Systems, *Man and Cybernetics* 2, vol. 21, pp. 445-452 (1991).

[9] L. Zadeh. Fuzzy Sets, *Information and Control*, Vol. 8, pp. 338-357 (1965).

[10] H.J. Zimmermann. *Fuzzy Sets: Theory and its Applications*. Kluwer Academic Publishers, Dordrecht (1985).

A genetic fuzzy classifier to adaptive user interest profiles with feature selection

María J. Martín-Bautista **María-Amparo Vila** **Henrik L. Larsen**

Dpt. of Computer Science and Artificial Intelligence
Granada University
Avda. Andalucía 37
18071 Granada, Spain
{mbautista, vila@decsai.ugr.es}

Department of Computer Science
Roskilde University, P.O. Box 260
DK-4000 Roskilde, Denmark
hll@ruc.dk

Abstract

We propose a method for guiding genetic algorithms for information retrieval by fuzzy classification and a genetic feature selection process of terms from documents evaluated by the user. The fuzzy classifier implements an inductive derivation of the current, experience based, interest profile in terms of an importance weighted conjunction of genes. A gene is defined by a symbol and a fuzzy number of occurrences of the symbol in documents belonging to the class of documents that satisfy the user's information need. Once the classification of the documents is made, a genetic selection from the most discriminatory terms is carried out. In this way, the terms that allow the system to discern between good and bad documents are selected and stored as a part of the user's profile to be used in future queries to the system. The fuzzy classification and term selection processes provide a better utilization of valuable knowledge for genetic algorithms in order to get an improvement of the quality of the estimates of the current and near future information needs in the areas of interest to the user.

Keywords: User profiles, Fuzzy Classification, Feature Selection, Genetic Algorithms, World Wide Web.

1 Introduction

The classification of documents in Web servers and document bases is a determinant aspect in Information Retrieval. The efficiency of the different methods utilized in query and matching processes is usually diminished by both a poor classification of the documents and a lack of personalization in the representation of the user's needs.

The classification problem has been widely studied in others disciplines such as Numerical Taxonomy and Machine Learning, and many techniques generated in these fields are being exported to solve the problems of classification and rules generation in Data Mining and Knowledge Discovery in all kinds of information systems, including Web and Textual ones.

The settling of the differences between the terms Web Mining and Text Mining comes from the aspects of the information access based on the user needs tasks versus the tasks of organization, classification and categorization of the textual information from the Web (Wulkefuhler & Punch, 1998) or other sources. Therefore, we understand by *Web Mining* the results of the application of traditional mine tasks to the information access and retrieving processes in relation to the web user needs.

When a user retrieves documents from an information system (for instance, from the Internet), most of actual systems have a lack of building a user profile. Hence, the user can not use the previous queries to the system for future requests. A learning of the user's needs is becoming a fundamental stage in the process of information retrieval. These needs may be represented by terms extracted from those documents that the user has evaluated as good ones.

One of the main problems studied in this field is the construction of user profiles by means of the discovering of the most relevant and representative terms (features) which information filtering systems can use to determine the most useful information to a given user. We must distinguish between those terms that best represent the information user needs, and those that allow us to discern between relevant and no-relevant, that is, the discriminatory terms for a certain classification.

2 Problem Formulation

Most of the techniques used in text classification are determined by the occurrences of the words (terms) appearing in the documents, combined with the user feedback over the documents retrieved. However, in our model, the most relevant terms will be selected from a previous fuzzy classification given by the genetic algorithm guided by the user feedback, but using techniques from Machine Learning.

Let $\Delta=\{D_1, \ldots, D_m\}$ be the set of documents evaluated by the user, and let $u_i \in [0, 1]$ be the user's evaluation of the document D_i, $i=1, \ldots, m$, meaning the degree to which the user finds that the document D_i satisfies his needs. We shall assume that an evaluation $u=0.5$ is neutral, while u>0.5 indicates a good document, $u=1$ representing a highly relevant document, while $u<0.5$ indicates a bad document, $u=0$ representing a document which is not relevant at all.

We will, without loss of generalization, assume that Δ is ordered decreasingly by u_i:

$$\Delta = \{D_1, \mathrm{K}, D_k, D_{k+1}, \mathrm{K}, D_{l-1}, D_l, \mathrm{K}, D_m\}$$

such that the subsets $\{D_1, \mathrm{K}, D_k\}$, $\{D_{k+1}, \mathrm{K}, D_{l-1}\}$, and $\{D_l, \mathrm{K}, D_m\}$ contains, respectively, the good, the neutral and the bad documents.

Let $T = \{t_1, \mathrm{K}\ t_n\}$ be the set of symbols extracted from Δ, and x_{ij} the relative frequency of symbol t_j in document D_i. The estimation of the expected value of x_j in good and bad documents is given by the weighed average of the relative occurrence frequency of a symbol t_j in the good and bad document by $\bar{x}_j$ and $\bar{x}'_j$, respectively:

$$\bar{x}_j = \frac{\sum_{i=1}^{k}(u_i \cdot x_{ij})}{\sum_{i=1}^{k} u_i} \qquad x'_j = \frac{\sum_{i=l}^{m}((1-u_i)\cdot x_{ij})}{\sum_{i=l}^{m}(1-u_i)} \qquad (1)$$

In our model, a gene in a chromosome is defined by a symbol $\mu_{\tilde{=}\eta}$ and a fuzzy number of occurrences of the symbol in documents belonging to the class of documents that satisfy the user's information need. Then G is a pair $G(t, \tilde{\eta})$, where t is a term, and $\tilde{\eta}$ is a fuzzy number characterized by the membership function $\mu_{\tilde{=}\eta}$ .as follows:

$$\mu_{\tilde{=}}(x) = \begin{cases} 0 & x = 0 \\ e^{-\frac{1}{2}\left(\frac{x-\eta}{\sigma}\right)} & x > 0 \end{cases} \qquad (2)$$

The membership function is a Gaussian one, assuming that the relative symbol occurrence frequency is normal distributed $N(\eta, \sigma^2)$, where the parameters η, $\sigma \in \mathrm{R}^+$, with $\eta = \bar{x}_j$ and σ_j^2 defined as follows:

$$\sigma_j^2 = \frac{\sum_{i=1}^{k}(u_i x_{ij} - x_j)^2}{\sum_{i=2}^{k} u_i} \qquad (3)$$

3 Description of the System

Two main modules can be distinguished in our system, namely the genetic feature selection and the fuzzy classifier (see Figure 1). We describe these modules in the following.

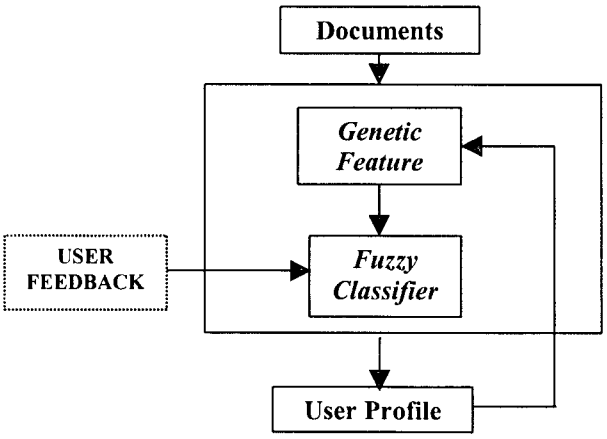


Figure 1. A view of the system.

3.1 The Genetic Selector Module

One of the first stages of the classification process is the Feature Selection (FS), by means of which the complexity of the problem is reduced by the elimination of irrelevant features to consider later in the classification stage.

The problem of FS in the framework of text databases, therefore, can be studied from two points of view:

1) If the documents are not previously classified, the selection of the most relevant features (terms in documents) give us those terms which describe the documents better.
2) On the other hand, if there is a previous classification (categorization) of the documents, a selection of the features would carry out the search of the most discriminatory terms, that is, those terms which allow to distinguish the different existent classes in a later stage. A study of the importance of the term reduction to improve some significant text categorization methods can be found in (Yang and Wilbur, 1996).

Given a previous classification, this module allows us to select the most discriminatory terms for a certain classification. In this case, we can select the discriminatory terms as derived from the fuzzy classification of the documents previously retrieved by the user. Through identifying and applying these terms, the system learns the user's interests, thus improving the quality of the fuzzy classification process when the user makes a new query.

The fitness function to maximize is based on the discriminatory power of every term through all the population as well as the accumulated discerning in the chromosomes.

$$p_y = s_y + q_y \qquad (4)$$

where s_y represents the similarity between a chromosome and the discriminatory vector, and it is based on the

Jaccard's score (Salton & McGill, 1983) weighted by an individual evaluation of every term:

$$S(C_y,V_h)=\frac{1}{|T|}\cdot\frac{\sum_{j=1}^{|T|}C_y(t_j)\cdot V_h(t_j)\cdot G(t_j)}{\sum_{j=1}^{|T|}C_y(t_j)+\sum_{j=1}^{|T|}V_h(t_j)-\sum_{j=1}^{|T|}C_y(t_j)\cdot V_h(t_j)} \quad (5)$$

where:

- $C_y(t_j)=\mu_{\Xi}(t_j)$, as defined in (2), and represents the relative frequency of every term t_j appearing in the gene of the chromosome C_y.
- $V_h(t_j)=\mu_{\Xi}(D_i(t_j))\cdot\mu_{\Xi}(D_k(t_j))$ is the discriminatory vector based on the comparision of the term t_i appearing in documents D_i and D_k.
- $G(t_j)=\frac{g(t_j)}{\sum_{j=1}^{|T|}g(t_j)}$, with $g(t_i)$ being the accumulated value of t_i for all the discriminatory vectors,

and q_y represents the capability of the chromosome itself, calculated by adding the accumulated discriminatory values of every term presented into the chromosome, as it is shown below.

$$q_y=\frac{\sum_{j=1}^{|T|}C_y(t_j)g(t_i)}{\sum_{j=1}^{|T|}g(t_j)} \quad (6)$$

3.2 The Fuzzy Classifier Module

The fuzzy classifier implements an inductive derivation of the current, experience based, interest profile in terms of an importance weighted conjunction of genes.

One of the most remarkable models of representation of documents is the vector space model, in which the terms of queries and documents are the components of vectors. These terms may be viewed as features which binary values 0 and 1 indicate the absence or presence of a term in the document, respectively. However, a more accepted representation of a document comes from a weighted vector, where every position indicates the term frequency, that is, the number of times that the term appears in the document, or the term importance indicator calculated by the product of the term frequency and the inverse document frequency, indicating the term frequency of a word in a document relative to the entire collection of documents (Salton, 1989).

A document score in a gene is been given by $\alpha=\mu_{\Xi\eta}(x)$, where x is the number of occurrences of the symbol t in the document.

Obviously, the aggregation of a document score in the genes of a chromosome query to its overall score for the chromosome query should apply an AND-like operator. On the other hand, the aggregation of the document-in chromosome scores in the whole population should be an OR-like operator. The selection of these operators for the production system should be based on their evaluation in an experimental setting.

Fitness Function

The evaluation of the Genetic Algorithm in this stage will be guided by the maximization of the fitness function, calculated from the combination of fuzzy precision and fuzzy recall.

The fuzzy recall-precision measure is applied in experimental situations where documents in the collection queried have all been evaluated by the user. Let $\Omega=\{\omega_1,\mathrm{K},\omega_{n_\Omega}\}$ be the collection queried, and let u_i and s_i be the (expert) user's and the system's evaluation of the document ω_i.

We define the fuzzy recall-precision τ by:

$$\tau=\rho^{v_1}\psi^{v_2} \quad (7)$$

where ρ is the fuzzy recall, and ψ is the fuzzy precision defined by:

$$\rho=\frac{\sum_{i=1}^{n_\Omega}min(u_i,s_i)}{\sum_{i=1}^{n_\Omega}u_i} \qquad \psi=\frac{\sum_{i=1}^{n_\Omega}min(u_i,s_i)}{\sum_{i=1}^{n_\Omega}s_i} \quad (8)$$

and v_1, v_2 are the importances of high recall and high precision, respectively.

For and Internet information retrieval system, we expect that precision is more important than recall, and therefore $v_1<v_2$.

Notice, that applied to the subset of Ω comprising the documents evaluated by the user in answer retrieved by the system, the fuzzy recall-precision τ measures how close the system's evaluation is to the user's evaluation.

4 Related Work

Several approaches using GAs related to this topic can be found in the literature. In BEAGLE (Ferguson, 1995). The author builds a population of user profiles which represent the best subset of keywords that allow us to discern among all the documents set those to be relevant.

There are some others approaches that use other techniques. Bloedorn et al. (Bloedorn, Mani & MacMillan, 1996), combine different learning methods such as Rocchio, C4.5 and AQ15, and measures coming from Information Retrieval and Machine Learning, to evaluate the influence of text features on user profiles, partitioning the document set into relevant and non-relevant ones. In (Pazzani and Billsus, 1997), a Bayesian classifier is used to define user profiles, and the expected information gain of the most informative terms is calculated as feature selection stage.

Collaborative filtering is seen as a classification task in (Pazzani and Billsus, 1998). The dimensionality of document terms is reduced by the selection of the most informative terms based on the singular value decomposition (SVD) of an initial matrix of user evaluations.

Other approach using GAs is presented in (Martín-Bautista, Larsen and Vila, 1998), where a user profile is built from the user preferences, represented by a population of chromosomes. Each chromosome is a vector of fuzzy genes, where every gene represents by a fuzzy set

the number of occurrences that a term should have in the document required by the user.

References

Baker, J.E. (1985) "Adaptive Selection Methods for Genetic Algorithms". In *Proc. on the First International Conference on Genetic Algorithms and their applications,* pp.101-111, Grefenstette, J.J. (ed). Hillsdale, New Jersey: Lawrence Earlbaum.

Bloedorn, F., Mani, I. & MacMillan, T.R. (1996). "Machine Learning of User Profiles: Representational Issues". In *Proceedings of AAAI'96,* . pp. 433-438. Portland, OR

Dash, M. & Liu, H. (1997). "Feature selection for Classification". In *Intelligent Data Analysis, vol 1, no.3.*

Ferguson, S. (1995). "BEAGLE: A genetic algorithm for Information Filter Profile Creation". *Technical Report CS-692.*University of Alabama.

Holland, J.H. (1992) *Adaptation in Natural and Artificial Systems*. Massachusetts: MIT Press.

John, G.H., Kohavi, R. and Pfleger, K. (1994) "Irrelevant Features and the Subset Selection Problem". In *Proc. of the Eleventh International Conference on Machine Learning,* pp.121-129. San Francisco, CA: Morgan Kauffmann Publishers.

Kraft, D.H., Petry, F.E., Buckles, B.P., & Sadasivan, T. (1995). Applying genetic algorithms to information retrieval systems via relevance feedback. In P. Bosc & J. Kacprzyk (Eds.), *Fuzziness in Database Management Systems* (pp. 330-344). Germany: Physica-Verlag.

Langley, P. (1994) "Selection of Relevant Features in Machine Learning". In *Proc. of the AAAI Fall Symposium on Relevance.* New Orleans, LA: AAAI Press.

Larsen, H.L., Andreasen, T.,& Christiansen, H. (1998) "Knowledge Discovery for Flexible Querying". In Christiansen, H., Andreasen, T. And Larsent, H.L. (Eds.) *Flexible Query Answering System.* Lecture Notes in Artificial Intelligence, vol. 1495, pp. 227-235. Berlin: Springer Verlag.

Martín-Bautista, M.J., Larsen, H.L. and Vila, M.A. (1999) A "Fuzzy Genetic Algorithm Approach to an Adaptive Information Retrieval Agent". *Journal of the American Society for Information Science* (Toappear).

Martín-Bautista, M.J. and Vila, M.A. (1998) "Applying Genetic Algorithms to the Feature Selection Problem in Information Retrieval". In *Lecture Notes On Artificial Intelligence (LNAI), 1495.* Springer-Verlag.

Martín-Bautista, M.J. and Vila, M.A. (1998) "A Survey of Genetic Feature Selection in Mining Issues". In *Proc. of Conference on Evolutionary Computation. (CEC'99).* July 1999, Washington. (To appear).

Mitchell, M. (1996) *An Introduction to Genetic Algorithms*. Cambridge, MA: MIT Press.

Pazzani, M. and Billsus, D. (1997) "Learning and Revising User Profiles: The identification of interesting web sites". *Machine Learning 27,* pp. 313-331.

Pazzani, M. and Billsus, D. (1998) "Learning Collaborative Information Filters". In *Proc. of the Fifteenth International Conference on Machine Learning.* San Francisco, CA: Morgan Kauffman Publishers.

Quinlan, J.R. (1992) *C4.5: Programs for Machine Learning.* Morgan Kaufmann.

Salton, G and McGill, M.J. (1983) *Introduction to Modern Information Retrieval.* New York: McGraw-Hill.

Salton, G. (1989) *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer.* Reading, PA: Addison-Wesley.

Siedlecki, W. and Sklansky, J. (1989) "A Note on Genetic Algorithms for Large-Scale Feature Selection". *Pattern Recognition Letters 10,* pp.335-347, North Holland: Elsevier Science Publishers.

Vafaie, H. and De Jong, K., (1992) "Genetic Algorithms as a Tool for Feature Selection in Machine Learning". In *Proceeding of the 4th International Conference on Tools with Artificial Intelligence,* Arlington, VA, November.

Wulfekulher, M.R. and Punch, W.F. (1998) "Finding Salient Features for Personal Web Page Categories". In *Hyper Proc. of the Sixth International World Wide Conference.*

Yang, J., Pai, P., Honavar, V. and Miller, L. (1998) "Mobile Intelligent Agents for Document Classification and Retrieval: A Machine Learning Approach". In *Proc. of the Fourteenth European Meeting on Cybernetics and Systems Research (EMCSR'98).* Vienna, Austria.

Yang, Y. and Wilbur, J. (1996) "Using Corpus Statistics to Remove Redundant Words in Text Categorization". *Journal of the American Society for Information Science, vol.47(5).*pp.357-369. John Wiley.

Zadeh, L.A. (1965). Fuzzy Sets. *Information and Control, 83,* 338-353.

Fuzzy classifier induction with GA-P algorithms

Luciano Sánchez Ramos
Universidad de Oviedo
Departamento de Informática
luciano@lsi.uniovi.es

Santiago García Carbajal
Universidad de Oviedo
Departamento de Informática
carbajal@lsi.uniovi.es

Summary

In this work we will use GA-P algorithms, a hybrid method between genetic algorithms and genetic programming, to induce fuzzy rule based classifiers with a more general structure than the attainable when applying GA's.

We will show that the numerical results this method gets are comparable to those produced by black-box classifiers.

Keywords: Genetic Programming, Genetic Fuzzy Systems, Fuzzy Classifiers.

1 Genetic Programming and GA-P Algorithms

Genetic programming has been recently regarded as the application of a genetic algorithm to optimize a function defined over the chains of a context-free language [2].

Crossover and mutation operations are specific to this problem. One or two point crossover of chains is not directly applicable, because the offspring would not always consist in valid chains for the language. The same thing can be said about mutation. Both cases are solved by codifying the chains with their parse trees or, as we will do here, by means of their syntax trees, annotated with the names of the production rules. If crossover is defined by the exchange of subtrees whose root is labeled with the same name, descendants are always parse or syntax trees of valid chains. It is remarkable that, under this point of view, GP crossover as defined by Koza [3] is a particular case of two-point GA crossover.

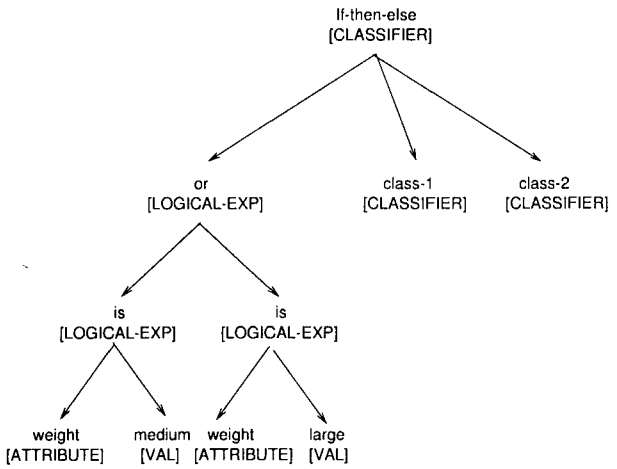


Figure 1: Annotated syntax tree of the chain "If weight is medium or weight is large then class-1 else class-2". The names of the non terminal symbols are shown in brackets.

1.1 Numerical coefficients evolution

If the grammar of the language contains numerical terminals, the values of these terminals are randomly assigned while generating the initial population and they will only evolve by mutation. As a practical consequence, final chains can be too complex (for instance, the number '2' could be represented by an expression like '(2.45+2.45)/2.45'). This is a problem with canonical GP.

Better results can be obtained if numbers are not terminal symbols, so mutation and crossover can operate on them and also when "named constants" are used. These are identifiers that are evaluated to a value and can be in more than one point of the chain.

An individual must contain two parts in order to use named constants: (1) the codification of their values and (2) the program that uses them. For instance, a polynomial that depends on a maximum of two different coefficients can be defined by means of the follow-

ing grammar:

S → C P

C → VALUE-K1 VALUE-K2

VALUE-K1 → D-CHAIN

VALUE-K2 → D-CHAIN

D-CHAIN → DIGIT | D-CHAIN DIGIT

DIGIT → 0|1|2|3|4|5|6|7|8|9

P → x_1 | x_2 | k_1 | k_2 | + P P | * P P

and the chain "123 11 * k_1 + k_1 * x_1 k_2" codifies the expression $k_1 * (k_1 + k_2 x_1) = 123 * (123 + 11\,x_1)$ in this language.

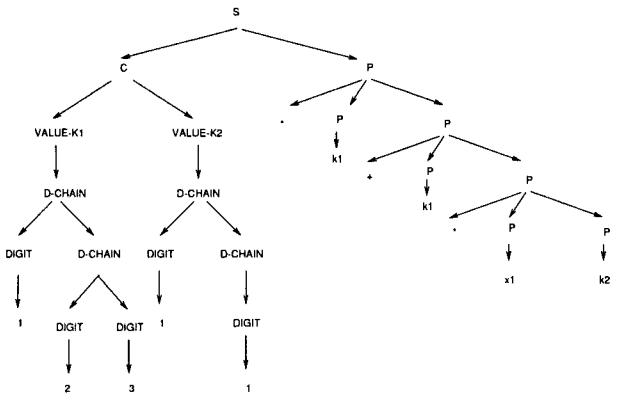


Figure 2: Parse tree of the chain "123 11 * k_1 + k_1 * x_1 k_2"

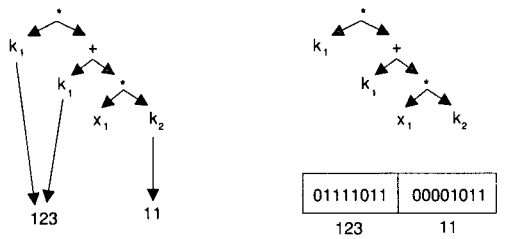


Figure 3: ADG of the chain "123 11 * k_1 + k_1 * x_1 k_2" and its GA-P representation

Representing a chain of the preceding language by means of its parse tree does not pose conceptual problems (see Figure 2) but it is very inefficient in time and space. We will use an acyclical graph best, that is analogous to the syntax tree of the chain (see figure 3).

A genetic algorithm that uses this last representation is named GA-P algorithm. In GA-P algorithms an individual comprises two parts: (a) the syntax tree of a chain of a language in which the identifiers of constants are terminal symbols and (b) the chain formed by the values of the constants, codified as in GA's (see Figure 3, right part). Mutation and crossover operations are defined to be equivalent to subtree exchange and random subtree replacement in the parse tree based representation.

2 Fuzzy rule based classifiers induction

A fuzzy rule based classifier can always be written as a chain in certain context-free language, so applying genetic programming to perform classifier induction is straightforward.

In this section we will use the names $C_1, C_2, \ldots, C_N$ to refer to the characteristics of an object (for example: C_1="length" and C_2="weight"). The name C_0 will be used to denote the variable "class", and this variable will take the values "class-1", "class-2", ..., "class-M". We will also use the names $V_{i1}, V_{i2}, \ldots, V_{in_i}$, $i = 0 \ldots N$, for denoting the symbolic labels of the n_i values of every variable. For instance, we can use V_{01} = "class-1", V_{02} = "class-2", V_{11} = "short", V_{12} = "medium", V_{13} = "long", and V_{21} = "light", V_{22} = "heavy".

2.1 Relation and rule based classifiers

We will define a fuzzy classifier by means of a fuzzy relation that maps every point in $C_0 \times C_1 \times \ldots \times C_N$ to a value between 0 and 1. Inference is made by intersecting the relation and the cylindrical extension of the fuzzy set that represents the measurement of the characteristics of the object being classified and projecting the intersection over C_0.

Every classifier defined by a relation can be written with rules, and the opposite is also true. Rule-based representation is useful to express linguistically the classifier's structure. We will use the following conventions:

2.1.1 Meaning of a rule

The following rule

$$\text{"if } C_i = V_{ij} \text{ then } \mathbf{t}(C_0 = V_{0k}) = \alpha\text{"}$$

where $\mathbf{t}(P)$ means "truth of assert P", is defined by the relation R_{ijk} that follows:

$$R_{ijk}(x_0, x_1, \ldots, x_N) = \begin{cases} \alpha & \text{If } x_i = V_{ij},\ x_0 = V_{0k} \\ 0 & \text{otherwise} \end{cases}$$

To shorten the notation, instead of writing

$$\text{"if } C_i = V_{ij} \text{ then } \mathbf{t}(C_0 = V_{0k}) = 1\text{"}.$$

we will write

$$\text{"if } C_i = V_{ij} \text{ then } C_0 = V_{0k}\text{"}$$

By last, the rule "if α then $C_0 = V_{0k}$" is defined by the relation $R_\alpha(V_{0k}, x_1, \cdots, x_N) = \alpha$, and $R_\alpha(x_0, x_1, \cdots, x_N) = 0$ for all $x_0 \neq V_{0k}$.

2.1.2 Logical expressions in antecedents

If the rules

$$\text{"if } A \text{ then } C_0 = V_{0k}\text{"}$$
$$\text{"if } B \text{ then } C_0 = V_{0k}\text{"}$$

are defined by means of the relations R_A and R_B, then the rules

$$\text{if } A \wedge B \text{ then } C_0 = V_{0k}$$
$$\text{if } A \vee B \text{ then } C_0 = V_{0k}$$

are defined by the relations

$$R_{A \wedge B}(x) = \min(R_A(x), R_B(x))$$
$$R_{A \vee B}(x) = \max(R_A(x), R_B(x))$$

respectively. With this definition, the rule

$$\text{"if } A \text{ then } \mathtt{t}(C_0 = V_{0k}) = \alpha\text{"}$$

is identical to this one:

$$\text{"if } A \wedge \alpha \text{ then } C_0 = V_{0k}\text{"}.$$

2.1.3 Rule concatenation

Concatenation of N rules defined by the relations $R_1, R_2, \ldots, R_N$ is defined by the relation $R(x) = \max\{R_1(x), \ldots, R_N(x)\}$. As a corollary, two rules with the same antecedent can be combined in one:

$$\text{if } A \text{ then } C_0 = V_{0k}$$
$$\text{if } B \text{ then } C_0 = V_{0k}$$

is identical to

$$\text{if } A \vee B \text{ then } C_0 = V_{0k}$$

2.2 Induction of the deep structure of the bank

The "deep structure" of the bank of fuzzy rules comprises rule definitions and the semantic associated to every value, while the definition of the classifier without specifying the fuzzy sets linked to the semantics of the variables is named "surface structure" after Zadeh [5]. Surface structure of a classifier has been induced with genetic programming if the memberships of the linguistic variables are known [2]. We propose a method for simultaneously evolving fuzzy partitions of the variables and rule definitions, that is inducting the deep structure of a bank from examples.

The first step consists in parameterizing a Ruspini's partition for every characteristic. We use triangular membership functions and decided that no measures are compatible with more than two linguistic values, so a fuzzy partition depends on so many parameters as linguistic values. Then, the semantics of a fuzzy classifier can be codified with a set of $n_1 + n_2 + \ldots + n_N$ numbers.

The second step consists in choosing an adequate grammar. We will use a grammar like the one that follows, and GA-P representation, to codify a complete fuzzy classifier:

CLASSIFIER →
V-LOGICAL-CT V-PARTITION-CT
if COND then $C_0 = V_{01}$
if COND then $C_0 = V_{02}$
...
if COND then $C_0 = V_{0M}$

COND → ASSERT-1 | ... | ASSERT-N |
(COND ∨ COND) | (COND ∧ COND) |
(COND ∧ LOGICAL-CT)

LOGICAL-CT → K_1 K_2 ... K_{N_K}

ASSERT-1 → trapezium-left(x_1,K_{11},K_{12}) |
triangle(x_1,K_{11},K_{12},K_{13}) | ...
trapezium-right(x_1,K_{1n_1-1},K_{1n_1})
...

ASSERT-N → trapezium-left(x_N,K_{N1},K_{N2}) |
triangle(x_N,K_{N1},K_{N2},K_{N3}) | ...
trapezium-right(x_N,K_{Nn_N-1},K_{Nn_N})

Non terminal symbol V-LOGICAL-CT is defined by concatenating the representation of the values of K_1 K_2 ... K_{N_K} and V-PARTITION-CT contains the values of K_{11} ... K_{1n_1} ... K_{N1} ... K_{Nn_N}

3 Numerical results

We have chosen the benchmark defined in [4] to contrast GA-P algorithms when inducting fuzzy classifiers. Every experiment has been repeated 30 times over three different permutations of the dataset. 75% of samples were used to design the classifier, 25% to

Dataset	K-NN	Linear Mean Dev	MLP Mean Dev	GA-P Mean Dev	Complexity Nodes	Variables Used Total
Cancer-1	1.7	2.93 0.18	1.38 0.49	3.06 1.00	37.7	4.5/9
Cancer-2	4.0	5.00 0.61	4.77 0.94	6.30 1.99	36.5	4.4/9
Cancer-3	4.5	5.17 0.00	3.70 0.52	4.50 1.79	35	4.4/9
Thyroid-1	5.95	6.56 0.00	2.38 0.35	4.92 0.94	34.5	4.1/21
Thyroid-2	6.00	6.56 0.00	1.91 0.24	4.80 0.82	37.7	4.5/21
Thyroid-3	6.50	7.23 0.02	2.27 0.32	4.88 0.88	36.6	4.3/21
Pima-1	25.5	25.83 0.56	24.10 1.91	24.92 1.89	34.1	3.7/8
Pima-2	27.6	24.69 0.61	26.42 2.26	29.68 2.05	24.3	3/8
Pima-3	23.5	22.92 0.35	22.59 2.23	24.29 2.17	25.4	3/8
Glass-1	35.8	46.04 2.21	32.70 5.34	39.62 4.73	73.9	6.7/9
Glass-2	33.9	55.28 1.27	55.57 3.70	42.11 2.78	81	5.8/9
Glass-3	35.0	60.57 3.82	58.40 7.82	41.15 4.72	68.5	5.3/9

Figure 4: Compared results between fuzzy classifiers obtained with GA-P algorithms and statistical and neural classifiers over the benchmark defined in [4].

test it by cross-validation. In Figure 4 the mean values obtained are shown for every permutation along with the standard deviations of the experiments. In all cases the algorithm was stopped after evaluating 50000 times the fitness function. Populations comprise 10 niches of 100 individuals each. The probability of crossover is 0.95, steady state was used and selection is performed with a tournament of size 4. Offspring replaces losers in the tournament.

Data relative to linear and neural classifiers is taken from [4]. The parameter of the k-neighbors classifier was chosen to minimize the classification error evaluated with jackknife. Observe that the dispersion of the results obtained when training a neural network from different starting points can be higher than the dispersion of GA-P results.

Since a classifier that decides between M classes can always be written with M rules, counting the number of rules makes no sense here. We have measured the complexity with the number of nodes in the classifier's syntax tree instead. Last column displays the mean number of characteristics a classifiers use.

4 Concluding Remarks

When designing fuzzy rule based classifiers, one of the main concerns is how much predictive accuracy has to be sacrificed for the sake of intelligibility. We have shown that GA-P algorithms can be used to induce the deep structure of fuzzy classifiers from data with good numerical behavior and obtaining compact solutions.

References

[1] Banzhaf, W. et al *Genetic Programming: An Introduction.* Morgan Kaufmann, 1998.

[2] Geyer-Schulz, A. *Fuzzy Rule-Based Expert Systems and Genetic Machine Learning.* Second edition. Physica-Verlag, 1997.

[3] Koza, J., *Genetic Programming.* MIT Press, Cambridge, MA. 1992

[4] Prechelt, Lutz. "PROBEN1 – A set of benchmarks and benchmarking rules for neural network training algorithms". Tech. Rep. 21/94, Fakultät für Informatik, Universität Karlsruhe, 1994.

[5] Zadeh, L. "Fuzzy Logic, Neural Networks and Soft Computing". Communications of the ACM, **37**(3), pp. 77-84. 1994.

LEARNING QUERIES FOR A FUZZY INFORMATION RETRIEVAL SYSTEM BY MEANS OF GA-P TECHNIQUES

Oscar Cordón
Dept. of Computer Science and A.I.
University of Granada
18071 - Granada (Spain)
e-mail: ocordon@decsai.ugr.es

Félix de Moya
Dept. of Librarianship
University of Granada
18071 - Granada (Spain)
e-mail: felix@goliat.ugr.es

Carmen Zarco
BARATZ - Servicios de Teledocumentación
Fuencarral, 123, 3º. Madrid (Spain)
e-mail: mcarmen@baratz.es

Summary

In this contribution, a Genetic Algorithm-Programming algorithm is proposed to automatically learn weighted queries —modeling the user's needs— for a Fuzzy Information Retrieval System. Although the fuzzy retrieval model constitute a powerful extension of the boolean one, the problem is that users are not usually able to express their query requirements in the form of a weighted query including the boolean operators AND, OR and NOT. The proposed model will be able to assist the user to obtain such queries by starting from a set of relevant documents and by applying an off-line adaptive process.

Keywords: Fuzzy Information Retrieval, GA-P algorithms, weighted queries, automatic query learning.

1 Introduction

Information Retrieval (IR) may be defined, in general, as the problem of the selection of documentary information from storage in response to search questions provided by an user [11]. Information Retrieval Systems (IRSs) are a kind of Information Systems that deal with data bases composed of information items —documents that may consist of textual, pictorial or vocal information— and process user queries trying to allow the user to access to relevant information in an appropiate time interval.

Most of the commercial IRSs are based on the boolean IR model [11], which presents some limitations. One of them is that boolean IRSs lack the ability to deal with imprecision and subjectivity, both of which are inherently present in the IR process. This is why the Fuzzy Information Retrieval (FIR) paradigm [3] has emerged as an active research area in the past decade.

FIRSs present a query language that allows the user to build queries composed of statements of positive or negative terms —numericaly or linguisticaly weighted— joined by boolean operators AND and OR. This query structure allows the system to improve the retrieval activity solving the boolean model problems.

Nevertheless, the use of weights requires a clear knowledge of the query semantics to translate a fuzzy concept into a crisp numeric value in [0,1]. On the other hand, both in boolean and FIRSs, it is difficult for non expert users to express their retrieval needs in the form of a query composed of different statements joined by the logical operators AND and OR. To solve these two problems and assist the user in the design of fuzzy queries, an automatic learning process based on the Genetic Algorithm-Programming (GA-P) paradigm [7] will be proposed in this paper.

In the following Sections, the basis of FIR and the composition of the proposed process are introduced. A preliminary experimentation will be also developed and some future research lines will be mentioned.

2 Fuzzy Information Retrieval

An IRS is constituted by three main components:

1. A *documentary data base*, which stores the documents and the representation of their information contents (usually based on index terms for textual documents). The *indexer module*, which automatically generates for each document its representation by extracting the document contents.

   In FIR, document representations and terms become fuzzy sets defined in the universe of terms and documents respectively, thus introducing a degree of relevance (aboutness) between a document and a term.

2. A *query subsystem*, which allows the users to formulate their queries and presents the relevant doc-

uments to them retrieved by the system. To do so, it includes a *query language*, that collects the rules to generate legitimate queries, and procedures to select the relevant documents.

FIRSs consider numeric or linguistic weigths in the query with different semantics [1], thus allowing the user to quantify the "subjective importance" of the selection requirements.

3. A *matching or evaluation mechanism*, which evaluates the degree to which the document representations satisfy the requirements expressed in the query and retrieves those documents that are judged to be relevant to it.

   A degree of document relevance is introduced in FIR —the so called *retrieval status value* (RSV)— by considering fuzzy operators appropriately performing the matching of queries to documents in a way that preserves the semantics of the former.

Let D be a set of documents and T be a set of unique and significant terms existing in them. The indexer module of the FIRS defines an indexing function:

$$F : D \times T \rightarrow [0,1]$$

In this paper, we use the normalized *inverted document frequency* [11]:

$$w_{d,t} = f_{d,t} \cdot log(N/N_t) \quad ; \quad F(d,t) = \frac{w_{d,t}}{Max_d\, w_{d,t}}$$

By projecting this "aboutness" fuzzy relation, a fuzzy set can be associated to each document and term:

$$d_i = \{< t, \mu_{d_i}(t) > | \, t \in T\} \quad ; \quad \mu_{d_i}(t) = F(d_i, t)$$

$$t_j = \{< d, \mu_{t_j}(d) > | \, d \in D\} \quad ; \quad \mu_{t_j}(d) = F(d, t_j)$$

where $f_{d,t}$ is the frequency of term t in document d, N is the number of documents in the collection and N_t is the number of documents containing term t.

Once the RSV of each document is known, the query subsystem affords a fuzzy set q specifying the degree of relevance of each document to the processed query:

$$q = \{< d, \mu_q(d) > | \, d \in D\} \quad ; \quad \mu_q(d) = RSV_q(d)$$

Thus, documents can be ranked in order of relevance in the query response. The user can specify the final relevant document set in two different ways: providing an upper bound for the number of retrieved documents or defining a threshold σ for the relevance degree.

The matching mechanism operates in a different way depending on the interpretation associated to the numeric weights included in the query (see [1, 3]). In this contribution, we consider the *importance* interpretation, where the weights represent the relative importance of each term in the query.

In this case, the RSV of each document to a fuzzy query q is computed as follows [12]. To maintain the semantics of the query, the term weighting has to take a different form according as the single term queries are ANDed or ORed. Therefore, assuming that A is a fuzzy term with assigned weight w, the following expressions are applied to obtain the fuzzy set associated to the weighted single term queries A_w (in the case of *disjunctive queries*) and A^w (for *conjunctive ones*):

$$A_w = \{< d, \mu_{A_w}(d) > | \, d \in D\}$$

$$\mu_{A_w}(d) = Min\ (w, \mu_A(d))$$

$$A^w = \{< d, \mu_{A^w}(d) > | \, d \in D\}$$

$$\mu_{A^w}(d) = Max\ (1 - w, \mu_A(d))$$

On the other hand, if the term is negated in the query, a negation function is applied:

$$\overline{A} = \{< d, \mu_{\overline{A}}(d) > | \, d \in D\} \quad ; \quad \mu_{\overline{A}}(d) = 1 - \mu_A(d)$$

The single fuzzy set representing the RSV of the compound query is obtained by combining the weighted term fuzzy sets by means of the following operators:

$$A\ AND\ B = \{< d, \mu_{A\ AND\ B}(d) > | \, d \in D\}$$

$$\mu_{A\ AND\ B}(d) = Min(\mu_A(d), \mu_B(d))$$

$$A\ OR\ B = \{< d, \mu_{A\ OR\ B}(d) > | \, d \in D\}$$

$$\mu_{A\ OR\ B}(d) = Max(\mu_A(d), \mu_B(d))$$

3 A GA-P Algorithm to Learn Fuzzy Queries

Genetic Algorithms (GAs) [10] have been succesfully applied as a tool to assist the user in the query formulation process (see [5] for a review including this and other different applications of GAs to IR). One of the approaches is based on automatically generating the query best describing the user's needs —represented in the form of an initial set of relevant (and optionally non relevant) documents— by means of an off-line process in which no user interaction is required. This operation mode is included in the usual *machine learning paradigm* and has been called *inductive query by example* (IQBE) by Chen in [2].

In [9], Kraft et al. proposed an automatic process of this kind to learn the whole composition of extended boolean queries (terms, weights and logical operators) for an FIRS. It is based on a variant of GAs, Genetic Programming (GP) [8], which evolves structures encoding programs such us expression trees. GP algorithms perform really well in the generation of structures, which are adapted both by crossover and mutation, but behaves bad in the learning of numerical

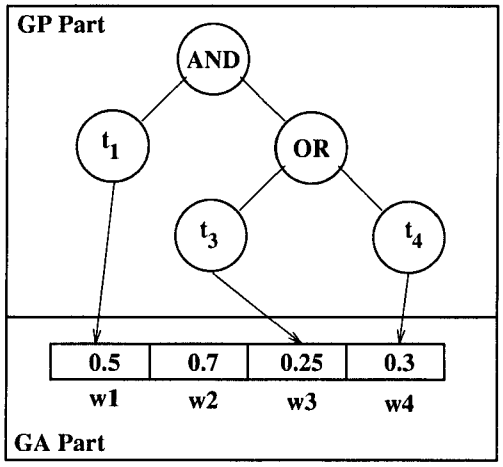


Figure 1: GA-P individual representing the fuzzy query 0.5 t_1 AND (0.25 t_3 OR 0.3 t_4)

values, which are only adapted by mutation. This is of significant importance in our case because these values corresponds to the query term weights.

The Genetic Algorithm-Programming (GA-P) [7], based on combining the traditional GAs with the GP technique to evolve complex expressions capable of handling numeric and symbolic data, can solve this problem. Each population member consists of both a value string and an expression as it is shown in Figure 3. The GP part of the GA-P evolves the expression (the query composition —terms and logical operators— in our case), while the GA part concurrently evolves the coefficients (the term weights). This is why our aim is to extend Kraft et al.'s algorithm to adequatelly deal with the term weights of the fuzzy query by considering a GA-P algorithm (for other applications of GA-P to Fuzzy Systems see [4, 13]).

The GA-P and GP make selection and child generation similarly, except that in the GA-P, crossover and mutation take place independently for the coefficient string and the expression component. Mutation and crossover rates for the coefficient string (using traditional GA methods) are independent from the rates for the expression part (using standard GP methods).

In our algorithm, the selection is steady-state [10]. Hence, each generation involves selecting two parents at random with a probability that grows with his adaption level —we consider the proportional scheme and the universal stochastic procedure—, crossing and mutating their GA and GP parts with probabilities $\{P_c^{GA}, P_m^{GA}\}$ and $\{P_c^{GP}, P_m^{GP}\}$ respectively, and introducing them in the current population only if they are better adapted than the worst two individuals.

To adapt the GP part, the usual GP crossover is considered [8], which randomly selects one edge in each parent and exchanges both subtrees from these edges between the both parents. Mutation is performed by a random selection of an edge and a random generation of a new subtree that substitutes the old one.

The GA part has a real coding issue and is adapted by the *Max-min-arithmetical crossover* [6] and the *Michalewicz's non-uniform mutation operator* [10]. Since the former works by generating four different offspring and selecting the best two ones, we combine the four GA parts with the two GP parts resulting from the GP crossover, thus obtaining two different groups of four descendants. The best from each group is the offspring finally considered.

The *generation of the initial population* is performed by computing random values in [0,1] for the GA part and random trees representing queries with a maximum predefined length and composed of randomly selected terms existing in the initial relevant documents provided by the user, for the GP part.

For the definition of the *fitness function*, two different possibilities are considered based on the classical precision and recall measures [9]:

$$F_1 = \frac{\sum_d r_d \cdot f_d}{\sum_d r_d} \quad ; \quad F_2 = \alpha \cdot \frac{\sum_d r_d \cdot f_d}{\sum_d r_d} + \beta \cdot \frac{\sum_d r_d \cdot f_d}{\sum_d f_d}$$

with $r_d \in \{0,1\}$ being the relevance of document d and $f_d \in \{0,1\}$ being the retrieval of document d in the processing of the current query. Simplest queries are preferred when having the same fitness value.

4 Experiments Developed

We have followed a similar experimental methodology to [9]. A documentary base composed of 359 abstracts taken from the *Library and Information Science Abstracts (LISA)* data base have been automatically indexed, obtaining a total number of 2609 different indexing terms. A user has selected a set of 82 relevant documents that has been provided to the system, which has been run considering the two different said fitness functions —with different values for α and β in F_2— and the use or not of the NOT operator.

The results obtained are shown in Tables 2 and 3, while the common parameter values are collected in Table 1. In the result tables, NOT stands for the consideration of the NOT operator, $\#rt$ for the number of documents retrieved by the learnt query, $\#rr$ for the number of relevant documents retrieved, R for recall, P for precision, and Sz for the generated query size.

In view of them, the process is not performing suitably. This fact is due to a main reason: *the composition of the fitness functions is not appropriated to represent*

Table 1: Common parameter values considered

Parameter	Decision
Population size	30
Number of generations	10000 (steady state)
GA Cross. probability	0.6
GA Mutation probability	0.1
GP Cross. probability	0.95
GP Mutation probability	0.05
Expression part limited to	15 nodes
Retrieval threshold σ	0.5

Table 2: Results obtained with fitness function F_1

NOT	$\#rt/\#rr$	R	P	Sz
N	359/82	1.0000	0.2284	3
Y	356/82	1.0000	0.2303	1

Table 3: Results obtained with fitness function F_2

(α,β)	NOT	$\#rt/\#rr$	R	P	Sz
(0.5,0.5)	N	359/82	1.0000	0.2284	3
(0.5,0.5)	Y	349/82	1.0000	0.2349	9
(0.5,1.0)	N	5/5	**0.0610**	**1.0000**	11
(0.5,1.0)	Y	1/1	0.0122	1.0000	1
(1.0,0.5)	N	359/82	1.0000	0.2284	3
(1.0,0.5)	Y	352/82	1.0000	0.2330	5
(0.5,0.75)	N	4/4	0.0488	1.0000	5
(0.5,0.75)	Y	1/1	0.0122	1.0000	1
(0.75,0.5)	N	359/82	1.0000	0.2284	3
(0.75,0.5)	Y	342/82	**1.0000**	**0.2398**	9

the problem goal. Focusing on F_1, we obtain two optimal solutions, both being very simple queries with total recall. Anyway, they are not useful due the large quantity of documents retrieved, goal not included in the fitness function guiding the search.

On the other hand, F_2 seemed to be able to solve this problem, because it considers two different criteria representing both problem goals. The problem is the weighted combination of the two criteria, which biases the search to one of the extremes of the pareto set of optimal solutions, thus causing a premature convergence of the GA-P algorithm. As seen in Table 3, solutions in an opposite part of the pareto are obtained when one of the weights dominates the other one, but the algorithm is not able to generate solutions with and appropriate balance between both criteria.

Multi-objective Evolutionary Algorithms can solve this problem since they look for compromise solutions satisfying all the criteria to different degrees. They also provide the user with a set of solutions differently distributed in the pareto, i.e., a set of candidate queries representing the user's needs in our problem. This will be our main research line in a very short future.

5 Concluding Remarks

A GA-P algorithm has been proposed to assist the user in the design of fuzzy queries for FIRSs. The process automatically generates complete extended boolean queries composed by terms, boolean operators and real-valued weights, describing the user's needs collected in a set of relevant documents provided by him.

The system did not behave properly in the experiments because of the wrong composition of the fitness functions, which biased the search to a specific zone of the pareto set. To solve this, multiobjective Evolutionary Algorithms will be considered in a future research.

References

[1] G. Bordogna, P. Carrara, G. Pasi, Fuzzy Approaches to Extend Boolean Information Retrieval, in: P. Bosc, J. Kacprzyk, Fuzziness in DataBase Management Systems (1995) 231-274.

[2] H. Chen, A Machine Learning Approach to Inductive Query by Examples: An Experiment Using Relevance Feedback, ID3, Genetic Algorithms, and Simulated Annealing, Journal of the American Society for Information Science 49:8 (1998) 693-705.

[3] V. Cross, Fuzzy Information Retrieval, Journal of Intelligent Information Systems 3 (1994) 29-56.

[4] O. Cordón, F. Herrera, L. Sánchez, Solving Electrical Distribution Problems Using Hybrid Evolutionary Data Analysis Techniques, Applied Intelligence 10:1 (1999) 5-24.

[5] O. Cordón, F. Moya, M.C. Zarco, A Brief Study on the Application of Genetic Algorithms to Information Retrieval (in spanish), Proc. 4th International Society for Knowledge Organization (ISKO) Conference (EOCONSID'99), Granada, Spain (april, 1999).

[6] F. Herrera, M. Lozano, J.L. Verdegay, Fuzzy Connectives Based Crossover Operators to Model Genetic Algorithms Population Diversity, Fuzzy Sets and Systems 92:1 (1997) 21-30.

[7] L. Howard, D. D'Angelo, The GA-P: a Genetic Algorithm and Genetic Programming Hybrid, IEEE Expert (1995) 11-15.

[8] J. Koza, Genetic Programming. On the Programming of Computers by Means of Natural Selection, The MIT Press (1992).

[9] D.H. Kraft, F.E. Petry, B.P. Buckles, T. Sadasivan, Genetic Algorithms for Query Optimization in Information Retrieval: Relevance Feedback, in: E. Sanchez, T. Shibata, L.A. Zadeh, Genetic Algorithms and Fuzzy Logic Systems (1997) 155-173.

[10] Z. Michalewicz, Genetic Algorithms + Data Structures = Evolution Programs, Springer-Verlag (1996).

[11] G. Salton, M.J. McGill, Introduction to Modern Information Retrieval, McGraw-Hill (1989).

[12] E. Sanchez, Importance in Knowledge Systems, Information Systems 14:6 (1989) 455-464.

[13] L. Sánchez, Interval-valued GA-P Algorithms, IEEE Trans. Evolutionary Computation (1999), to appear.

AGGREGATION FUNCTIONS AND FUZZY MEASURES: THE MULTI-DIMENSIONAL CASE

M. Mas
Dept. of Math. and Comp. Sc.
Univ. of the Balearic Islands
07071-Palma de Mallorca.Spain
dmimmg0@ps.uib.es

G. Mayor
Dept. of Math. and Comp. Sc.
Univ. of the Balearic Islands
07071-Palma de Mallorca.Spain
dmigmf0@ps.uib.es

J. Suñer
Dept. of Math. and Comp. Sc.
Univ. of the Balearic Islands
07071-Palma de Mallorca.Spain
dmijsl0@ps.uib.es

Summary

This paper is devoted to the study of multi-dimensional fuzzy measures and integrals from the point of view of aggregation operators. A necessary and sufficient condition for a multi-dimensional discrete Choquet (and Sugeno) integral to be a multi-dimensional aggregation function (MAF) is given. The particular case of fuzzy measures generated by a t-conorm is specially studied.

Keywords: Multi-dimensional aggregation function, fuzzy measure, discrete Sugeno and Choquet integrals, t-conorm, iteration.

1 INTRODUCTION

Fuzzy measures were introduced by M. Sugeno in 1974 in order to express a grade of fuzziness in the same way that probability measures express a grade of randomness. The Sugeno fuzzy integrals are the functionals with monotonicity defined by using fuzzy measures. Later on, Murofushi and Sugeno proposed another type of fuzzy integral, the Choquet integral, based on the Capacity Theory developed by G. Choquet in 1953. Many authors have investigated on the characterization of Sugeno and Choquet integrals (P. Wakker, D. Schmeidler, D. Dubois, H. Prade, R. Sabbadin, J. L. Marichal). Both fuzzy integrals have proved to be very useful as aggregation operators.

In this paper, we deal with the multi-dimensional aspect of fuzzy measures and fuzzy integrals, and we give conditions for such a multi-dimensional functions to be a Multi-dimensional Aggregation Function. This kind of aggregation functions have been already studied by G. Mayor and T. Calvo in [5].

2 PRELIMINARIES

In this section we present the definitions and basic results we will use throughout this paper.

Let us denote by E the set $E = \bigcup_{n\geq 1} [0,1]^n$ of all the ordered lists formed by elements of $[0,1]$.

Definition 1 *Given* $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_m)$ *elements of* E, *we define the following relations:*

- $x \leq_\pi y$ *if and only if* $n = m$ *and* $x_1 \leq y_1, \ldots, x_n \leq y_n$ *(product order)*
- $x \leq_\alpha y$ *if and only if* $n \leq m$, *and* $(x_1, \ldots, x_n, \overbrace{\vee x_i, \ldots, \vee x_i}^{m-n}) \leq_\pi (y_1, \ldots, y_m)$
- $x \leq_\beta y$ *if and only if* $n \geq m$, *and* $(x_1, \ldots, x_n) \leq_\pi (y_1, \ldots, y_m, \overbrace{\wedge y_i, \ldots, \wedge y_i}^{n-m})$

where $\vee x_i$ *and* $\wedge y_i$ *stand for* $\max\{x_1, \ldots, x_n\}$ *and* $\min\{y_1, \ldots, y_m\}$, *respectively.*

Proposition 1 *The relations* $\leq_\pi, \leq_\alpha$ *and* $\leq_\beta$ *are orders on the set* E.

Remark 1 a) $x \leq_\pi y$ if and only if $x \leq_\alpha y$ and $x \leq_\beta y$.

b) Given $x = (x_1, \ldots, x_n) \in E$, we will denote $\underline{x} = (x_1, \ldots, x_n, \wedge x_i)$ and $\overline{x} = (x_1, \ldots, x_n, \vee x_i)$. Observe that $\underline{x} \leq_\beta x \leq_\alpha \overline{x}$.

Consider a multi-dimensional function $f : E \to [0,1]$. Observe that it can be viewed as a sequence $f = f_1, f_2, \ldots$ where, for each $n \geq 1$, $f_n : [0,1]^n \to [0,1]$ is the restriction of f to $[0,1]^n$.

Definition 2 *A function* $f : E \to [0,1]$, $f = f_1, f_2, \ldots$, *is a Multi-dimensional Aggregation Function (f is a MAF) if it satisfies the following conditions:*

1) *Each* f_n *is continuous, non-decreasing and idempotent.*

2) *f is non-decreasing with respect to* $\leq_\alpha$ *and* $\leq_\beta$*:*

   a) $x \leq_\alpha y \Longrightarrow f(x) \leq f(y)$

   b) $x \leq_\beta y \Longrightarrow f(x) \leq f(y)$

A useful characterization of MAFs is the following:

Proposition 2 *A multi-dimensional function* $f = f_1, f_2, \ldots$ *is a MAF if and only if*

1) $f_1 = id$.

2) *Each* f_n *is continuous and non-decreasing.*

3) *For each* $n \geq 1$ *and for each* $x \in [0,1]^n$,
$$f_{n+1}(\underline{x}) \leq f_n(x) \leq f_{n+1}(\overline{x}).$$

Example 1 The following elementary multi-dimensional functions are MAFs:

a) $f_n(x_1, \ldots, x_n) = \min\{x_1, \ldots, x_n\}$,
$f_n(x_1, \ldots, x_n) = \max\{x_1, \ldots, x_n\}$.

b) $f_n(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}$,
$f_n(x_1, \ldots, x_n) = \sqrt[n]{x_1 \cdots x_n}$.

c) $f_n(x_1, \ldots, x_n) = x_1$, $f_n(x_1, \ldots, x_n) = x_n$.

Remark 2 Given a MAF f, then, for each $n \geq 1$ and $x \in [0,1]^n$,

$$\min\{x_1, \ldots, x_n\} \leq f(x_1, \ldots, x_n) \leq \max\{x_1, \ldots, x_n\}.$$

3 MULTI-DIMENSIONAL AGGREGATION FUNCTIONS GENERATED BY ITERATION

In this section we present a natural procedure of generating multi-dimensional aggregation functions from a bi-dimensional function.

Given a function $\varphi : [0,1]^2 \to [0,1]$, we can consider the multi-dimensional function $f = f_1, f_2, \ldots$ defined as:

1) $f_1 = id$, and

2) $f_n(x_1, \ldots, x_n) = \varphi(f_{n-1}(x_1, \ldots, x_{n-1}), x_n)$, $\forall\, n \geq 2$.

When f is generated by φ, we will write $f =< \varphi >$.

It is interesting to observe that when $\varphi = T$ is a t-norm, then $f =< \varphi >$ is monotonic with respect to $\leq_\beta$ and when $\varphi = S$ is a t-conorm, then $f =< \varphi >$ is monotonic with respect to $\leq_\alpha$. Moreover, if f is generated by a t-norm T and it is non decreasing with respect to $\leq_\alpha$, then $T = \min$. Analogously, if f is generated by a t-conorm S and it is non decreasing with respect to $\leq_\beta$, then $S = \max$.

Proposition 3 *Given* $\varphi : [0,1]^2 \to [0,1]$, *let us consider the multi-dimensional function* $f =< \varphi >$. *Then f is a MAF if and only if* φ *is continuous, non-decreasing and idempotent.*

Example 2 a) The MAFs of the example 1a are generated, respectively, by $\varphi = \min$ and $\varphi = \max$.

b) The arithmetic and geometric means are not generated by iteration. In fact, the only MAF $f =< \varphi >$ with each f_n a quasi-linear arithmetic mean has the form:

$$\begin{aligned} f_n(x_1, \ldots, x_n) &= g^{-1}((1-\lambda)^{n-1} g(x_1) + \\ &+ \lambda(1-\lambda)^{n-2} g(x_2) + \cdots + \\ &+ \lambda(1-\lambda) g(x_{n-1}) + \lambda g(x_n)) \end{aligned}$$

where $\lambda \in [0,1]$ and g is a continuous, strictly monotonic function defined on $[0,1]$.

c) The functions of the example 1c can be obtained by taking $\lambda = 0$ or 1 and $g = id$ in the expression above.

Proposition 4 *Given a MAF* $f =< \varphi >$, *then f is symmetric (each* f_n *is symmetric) if and only if there exists* $k \in [0,1]$ *such that*

$$f_n(x_1, \ldots, x_n) = \begin{cases} \max\{x_1, \ldots, x_n\} & \text{if } x_i \leq k\ \forall i \\ \min\{x_1, \ldots, x_n\} & \text{if } x_i \geq k\ \forall i \\ k & \text{otherwise} \end{cases}$$

for all n.

The proof of this proposition relies on the fact that if f is symmetric, then φ is associative and symmetric and hence, it is a (Max-Min)-operator, in the sense given in [4].

4 MULTI-DIMENSIONAL FUZZY MEASURES

For each $n \in \mathbb{N}$, let us consider the set $\mathbb{N}_n = \{1, 2, \ldots, n\}$.

Definition 3 *A fuzzy measure on* $\mathbb{N}_n$ *is a set function* $\mu_n : 2^{\mathbb{N}_n} \to [0,1]$ *satisfying the following conditions:*

1) $\mu_n(\emptyset) = 0,\ \mu_n(\mathbb{N}_n) = 1.$

2) $A \subset B \Rightarrow \mu_n(A) \leq \mu_n(B)\ \ \forall A, B \in 2^{\mathbb{N}_n}.$

From now on, given n numbers $x_1, \ldots, x_n$, $x_{(1)}, \ldots, x_{(n)}$ will denote the sequence arranged in increasing order: $x_{(1)} \leq \cdots \leq x_{(n)}$.

Definition 4 *Given a fuzzy measure μ, the n-dimensional discrete Sugeno integral with respect to μ is the function $S_\mu : [0,1]^n \to [0,1]$ defined by*

$$S_\mu(x_1, \ldots, x_n) = \bigvee_{i=1}^{n} (x_{(i)} \wedge \mu(A_{(i)}))$$

where $A_{(i)} = \{(i), (i+1), \ldots, (n)\}$.

Definition 5 *Given a fuzzy measure μ, the n-dimensional discrete Choquet integral with respect to μ is the function $C_\mu : [0,1]^n \to [0,1]$ defined by*

$$C_\mu(x_1, \ldots, x_n) = \sum_{i=1}^{n} (x_{(i)} - x_{(i-1)})\, \mu(A_{(i)})$$

with the same notations as above, and $x_{(0)} = 0$.

Definition 6 *A function $f : [0,1]^n \to [0,1]$ is called a discrete Sugeno (Choquet) integral if there exists a fuzzy measure μ on $\mathbb{N}_n$ such that $f = S_\mu$ ($f = C_\mu$).*

Remark 3 The Sugeno and Choquet integrals are idempotent, non-decreasing and continuous operators.

Different authors ([5]) have studied the problem of when a multi-dimensional OWA operator, or a multi-dimensional quasi-arithmetic weighted mean, $f = f_1, f_2, \ldots$ is a MAF. As we know ([2]), both the OWA operators ([7]) and the arithmetic means are discrete Choquet integrals (with respect to symmetric and additive fuzzy measures, respectively). Thus we propose now the problem of when a multi-dimensional discrete Choquet (or Sugeno) integral is a MAF. First of all, we must define such multi-dimensional functions.

Let us consider a sequence $\mu = \{\mu_n : n \geq 1\}$ of fuzzy measures, where each μ_n is defined on $\mathbb{N}_n$. Then we can consider the following two kinds of multi-dimensional operator $f : \bigcup_{n \geq 1} [0,1]^n \to [0,1]$:

- Multi-dimensional discrete Choquet integral: For each $n \geq 1$ and $(x_1, \ldots, x_n) \in [0,1]^n$, $f_n(x_1, \ldots, x_n) = C_{\mu_n}(x_1, \ldots, x_n)$.
- Multi-dimensional discrete Sugeno integral: For each $n \geq 1$ and $(x_1, \ldots, x_n) \in [0,1]^n$, $f_n(x_1, \ldots, x_n) = S_{\mu_n}(x_1, \ldots, x_n)$.

We would like to know when these operators are MAF. We already know that, for each $n \geq 1$, C_{μ_n} and S_{μ_n} are continuous, idempotent and non-decreasing. Then we must impose the non decreasingness with respect to the orders $\leq_\alpha$ and $\leq_\beta$.

Proposition 5 *Given $\mu = \mu_1, \mu_2, \ldots$ a sequence of fuzzy measures, let us consider the multi-dimensional function $f = f_1, f_2, \ldots$, where each f_n is a discrete Choquet integral with respect to μ_n. Then f is a MAF if and only if the sequence μ of fuzzy measures satisfies the inequalities*

$$\mu_{n+1}(A) \leq \mu_n(A) \leq \mu_{n+1}(A \cup \{n+1\})$$

$\forall\, n \geq 1$ and $A \in 2^{\mathbb{N}_n}$.

Proof. Suppose first that these two inequalities are satisfied. We prove the three conditions stated in the proposition 2. The first two conditions are obvious from the remark 4 and the third condition comes directly from the inequalities on μ. The proof of the converse is obtained by giving values 0 and 1 properly to the arguments of f_n and f_{n+1} in the condition 3 of the proposition 2.

It is interesting to observe that, when f is a multi-dimensional discrete Sugeno integral, we obtain the same result.

Remark 4 Observe that, when μ is a sequence of symmetric fuzzy measures (that is, f is a multi-dimensional OWA operator), this condition is equivalent to say that the weighting triangle $\bigwedge \omega_i^n$ associated to f is regular ([5]). An interesting study related to the regularity can be found in [1].

In the case of simple fuzzy measures, that is, $\mu(A) = 0$ or $1\ \ \forall\, A \in 2^{\mathbb{N}}$, the inequalities of proposition 5 are equivalent to the two conditions: $\forall n$, $\forall A \subset \{1, \ldots, n\}$,

$$\text{If } \mu_n(A) = 0 \Longrightarrow \mu_{n+1}(A) = 0$$

$$\text{If } \mu_n(A) = 1 \Longrightarrow \mu_{n+1}(A \cup \{n+1\}) = 1$$

5 MULTI-DIMENSIONAL FUZZY MEASURE GENERATED BY A T-CONORM

Let S be a t-conorm and, for each $n \geq 2$, $\alpha = (\alpha_1, \ldots, \alpha_n) \in [0,1]^n$ such that $S(\alpha_1, \ldots, \alpha_n) = 1$. We can define a fuzzy measure $\mu_{S,\alpha} : 2^{\mathbb{N}_n} \to [0,1]$ as

$$\mu_{S,\alpha}(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \alpha_i & \text{if } A = \{i\} \\ S(\alpha_{i_1}, \ldots, \alpha_{i_p}) & \text{if } A = \{i_1, \ldots, i_p\}, \\ & 2 \leq p \leq n \end{cases}$$

It is easy to proof the following result:

Proposition 6 *$\mu_{S,\alpha}$ is a fuzzy measure and, moreover,*

$$\max\{\mu_{S,\alpha}(A), \mu_{S,\alpha}(B)\} \leq \mu_{S,\alpha}(A \cup B) \leq S(\mu_{S,\alpha}(A), \mu_{S,\alpha}(B)),$$

$\forall A, B \in 2^{\mathbb{N}_n}$.

In the second inequality, the equality holds, $\mu_{S,\alpha}(A \cup B) = S(\mu_{S,\alpha}(A), \mu_{S,\alpha}(B)) \quad \forall A, B \in 2^{\mathbb{N}_n}$, if and only if $\alpha_1, \ldots, \alpha_n$ are S-idempotent. In this case, $\mu_{S,\alpha}(\{i_1, \ldots, i_p\}) = \max\{\alpha_{i_1}, \ldots, \alpha_{i_p}\}$, $2 \leq p \leq n$.

These fuzzy measures have a strong multi-dimensional character. It is enough, in the definition, to consider a triangle of α's, $\bigwedge \alpha_i^n$, such that $\alpha_1^1 = 1$ and the t-conorm applied to the elements of each row is 1: $S(\alpha_1^n, \ldots, \alpha_n^n) = 1$, $\forall n \geq 2$. Then, for each $n \geq 1$, μ_n is the fuzzy measure defined as above. We call μ the sequence $\{\mu_n : n \geq 1\}$. Next we give a necessary condition and a sufficient condition for such a fuzzy measure to produce a multi-dimensional aggregation function.

Proposition 7 *a) In the conditions above, if C_μ is a multi-dimensional aggregation function, then $\alpha_i^{n+1} \leq \alpha_i^n \leq S(\alpha_i^{n+1}, \alpha_{n+1}^{n+1})$ $\forall n \geq 1$, $\forall i = 1, \ldots, n$.*

b) Suppose that the triangle $\bigwedge \alpha_i^n$ satisfies that $\alpha_i^{n+1} \leq \alpha_i^n \leq S(\alpha_i^{n+1}, \alpha_{n+1}^{n+1})$ $\forall n \geq 1$, $\forall i = 1, \ldots, n$ and the elements α_n^n are S-idempotent for all $n \geq 1$. Then C_μ is a multi-dimensional aggregation function.

Remark 5 The converse is not true, in general, in both results. In effect, in the first case, consider the t-conorm $S(x, y) = \min\{x+y, 1\}$ and the triangle with third row $(\frac{1}{3}, \frac{1}{2}, \frac{1}{3})$ and fourth row $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, \frac{1}{8})$. Similarly, in the second case, consider the same t-conorm and the triangle with third row $(\frac{1}{3}, \frac{1}{2}, \frac{1}{3})$ and fourth row $(\frac{1}{4}, \frac{1}{3}, \frac{1}{4}, \frac{1}{3})$.

Observe that the condition of the elements α_n^n, $n \geq 1$, being S-idempotents is very strong. A usual multi-dimensional aggregation function like the arithmetic mean $f(x_1, \ldots, x_n) = \frac{x_1 + \cdots + x_n}{n}$ comes from the fuzzy measure $\mu(A) = \frac{|A|}{n}$, that is, $S(x, y) = \min\{x + y, 1\}$ and $\alpha_i^n = \frac{1}{n}$ $\forall i = 1, \ldots, n$. We give next an equivalent condition for the case of additive fuzzy measures which does not impose idempotency of any element of the triangle.

We know that, if μ_n is an additive fuzzy measure on $\mathbb{N}_n$, then the discrete Choquet integral with respect to μ_n, C_{μ_n}, is a weighted arithmetic mean, with weights $\omega_i^n = \mu_n(\{i\})$. We also know that such a measure must be generated by the t-conorm $S(x, y) = \min\{x + y, 1\}$ and $(\alpha_1, \ldots, \alpha_n)$ such that $\sum_{i=1}^{n} \alpha_i = 1$.

Proposition 8 *Let us consider a sequence μ such that every fuzzy measure μ_n $(n \geq 1)$ is additive. Then C_μ is a multi-dimensional aggregation function if and only if the triangle $\bigwedge \alpha_i^n$ satisfies that $\alpha_i^{n+1} \leq \alpha_i^n$ $\forall n \geq 1$, $\forall i = 1, \ldots, n$.*

Acknowledgements

This paper has been partially financed by CICYT, project: TIC96-1393-C06-06.

References

[1] M. Carbonell, M. Mas, G. Mayor. On a class of Monotonic Extended OWA Operators. *FUZZ-IEEE'97*, pp. 1695-1700. Barcelona, 1997.

[2] M. Grabisch. Fuzzy integral in multicriteria decision making. *Fuzzy Sets and Systems, 69*, pp. 279-298, 1995.

[3] M. Grabisch, H. Nguyen, E. Walker. Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference. *Kluwer Academic*, 1995.

[4] M. Mas, G. Mayor, J. Torrens. t-operators. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems.* Vol. 7, 1, pp 31-50, 1999.

[5] G. Mayor, T. Calvo. On Extended Aggregation Functions. *IFSA '97*, pp. 281-285. Praga, 1997.

[6] M. Sugeno. Theory of fuzzy integrals and its applications. PhD thesis, Tokyo Institute of Technology, 1974.

[7] R.R. Yager. On ordered weighted averaging aggregation operators in multi-criteria decision making. *IEEE Trans. on SMC, 18*, pp. 183-190, 1988.

GENERALISED QUALITATIVE UTILITY FUNCTIONS FOR REPRESENTING PARTIAL PREFERENCES RELATIONS

Lluís Godo **Adriana Zapico**

Institut d'Investigació en Intel.ligència Artificial - CSIC
Campus UAB s/n.
08193 Bellaterra, Spain.
{godo,zapico }@iiia.csic.es

Summary

Representational issues of preferences in the framework of a possibilistic (ordinal) decision model under uncertainty were introduced by Dubois and Prade quite recently. In this framework, (finite) linear uncertainty and preference scales are assumed, and decisions can be ranked according to their expected utility in terms of Sugeno integrals. In this paper we generalise the model by allowing (i) to measure uncertainty and preferences on non linear lattices, and (ii) to formulate the utility expectations in terms of generalized Sugeno integrals where t-norms and t-conorms play a role. For these generalised utility functions we provide axiomatic characterisations. Finally, we propose how to extend the utility functions to cope with belief states that may be partially inconsistent.

1 Introduction

The representational issues of preferences in the framework of a possibilistic (qualitative / ordinal) decision model under uncertainty, were originally introduced a few years ago by Dubois and Prade [4], and more recently linked to case-based decision problem in [3]. They proposed a qualitative counterpart to Von Neumann and Morgenstern's Expected Utility Theory [5]. In this approach, the uncertainty is assumed to be of possibilistic nature, i.e. belief states are represented by normalised possibility distributions π, and make use of finite commensurate ordinal preference and uncertainty scales. However, sometimes we may be faced with case-based decision problems where similarity may take values that are incomparable, Decision Maker's preferences may be partial, etc.. In many applications we have to measure degrees of similarity or preferences in partially ordered sets, so this may be an interesting subject which we will investigate in a possibilistic context. Hence, to cope with these situations, as a first step, an extension of the above approach is proposed in [6], where both preferences and uncertainty are graded on finite distributive lattices, $(U, \wedge, \vee, 0, 1, n_U)$ and $(V, \wedge, \vee, 0, 1, n_V)$ respectively, that are commensurate. Axiomatic settings for characterising two qualitative utilities, a pessimistic and an optimistic one are provided. Given $u : X \to U$ a preference function that assigns to each consequence of X a preference level of U, and $h : V \to U$ a $\{0,1\}$-homomorphism relating both lattices V and U. Let n be the reversing homomorphism $n : V \to U$ defined as $n(\lambda) = n_U(h(\lambda))$, also verifying $n(0) = 1$, $n(1) = 0$. For any normalised π, i.e. $\pi : X \to V$ s.t. exists $x|\ \pi(x) = 1$, we considered the qualitative utility functions:

$$QU^-(\pi|u) = \bigwedge_{x \in X}(n(\pi(x)) \vee u(x)),$$
$$QU^+(\pi|u) = \bigvee_{x \in X}(h(\pi(x)) \wedge u(x)).$$

QU^- is a pessimistic criterion that looks for those π's which, at some extent, make hardly plausible all the bad consequences. On the contrary, QU^+ is an optimistic criterion which looks for those π's that, also to some extent, makes plausible some of the good consequences.

In this paper we consider other alternatives to obtain generalised utility functions GQU^- and GQU^+, namely the use of a t-norm instead of $\wedge$ in QU^- and a t-conorm instead of $\vee$ in QU^+. We will show that, under some conditions, GQU^- and GQU^+ are indeed "utility" functions in the set of normalised possibility distributions, in the sense that they preserve the preference ordering and the "natural operation" of possibilistic mixture used to combine possibilistic lotteries or distributions. In a recent paper [7], we have proposed an extension of the possibilistic decision model to deal with non-normalised possibility distributions i.e. distributions accounting for partially belief states. Following this proposal, now we also take into account this ability in the generalised model and show its usefulness to provide elements for a qualitative case-based decision methodology.

In the following section we provide a background on lattices, while in section 3 the context of our work is given. In section 4 we propose two axiomatic settings for characterising both pessimistic and optimistic generalised qualitative utilities requiring finite distributive and commensurate lattices for assessing uncertainty

and preferences; also including an extension that lets us make decisions in contexts in which possibly partially inconsistent belief states are involved.

2 Lattices and Possibility Distributions

In this preliminary section we introduce some notation, definitions and results related to lattices (for more details see [2]) and possibility distributions that we will use in the rest of the paper.

Let L be a set and $\wedge$ and $\vee$ two binary operations on L. $(L, \wedge, \vee)$ is a *lattice* if $\wedge, \vee$ are associative, commutative, satisfy idempotency and the absorption laws. The induced order in the lattice is: $x \leq y$ iff $x \wedge y = x$. $(L, \wedge, \vee, n_L, 0, 1)$ will denote a *bounded lattice with involution*, i.e. $0, 1 \in L$, $0 \leq x \leq 1$, for all $x \in L$, and $n_L : L \to L$ a decreasing function s.t. $n_L(n_L(x)) = x$. Note that n_L satisfies that $n_L(0) = 1, n_L(1) = 0, n_L(x \wedge y) = n_L(x) \vee n_L(y)$ and $n_L(x \vee y) = n_L(x) \wedge n_L(y)$.

Given a partially pre-ordered set $(L, \leq)$, i.e. $\leq$ is reflexive and transitive, the associated indifference and incomparability relations are defined as:

$$x \sim y \iff x \leq y \text{ and } y \leq x.$$
$$x <> y \iff x \not\leq y \text{ and } y \not\leq x.$$

Let $L/\sim$ denote *the quotient set* w.r.t. the equivalence relation $\sim$ and let us denote also by $\leq$ the ordering induced by L into $L/\sim$. Then, we shall call $(L, \leq)$ a *pre-lattice* iff $(L/\sim, \leq)$ is a bounded lattice with the meet and joint operations defined from $\leq$. Now, given a lattice $(L, \wedge, \vee, 0, 1)$, *a t-norm* (t-conorm) operation $\top(\perp)$ on L is any non-decreasing, associative and commutative binary operation on L, verifying $\lambda \top 0 = 0$ and $\lambda \top 1 = \lambda$ ($\lambda \perp 0 = \lambda$ and $\lambda \perp 1 = 1$, resp.) for all $\lambda \in L$. The *residuum of* $\top$ is defined as

$$\mathrm{I(a, c)} = \bigvee\{b \in L \mid \top(\mathrm{a}, \mathrm{b}) \leq \mathrm{c}\}.$$

$(L, \wedge, \vee, \top, \mathrm{I}, 0, 1)$ is a *residuated lattice* if $(L, \wedge, \vee, 0, 1)$ is a lattice and $(\top, \mathrm{I})$ is an adjoint pair, i.e. if

- $(L, \top, 1)$ is a commutative semigroup with unit element 1.
- $\forall a, b, c \in L$ $(a \top b) \leq c$ iff $a \leq \mathrm{I}(b, c)$.

Finally, one can check that if we have a finite lattice $(L, \wedge, \vee, 0, 1)$ and a t-norm $\top$ on L, then $\top$ distributes over the lattice joint operation (i.e. $(a \vee b) \top c = (a \top c) \vee (b \top c), \forall a, b, c \in L$) iff $(L, \wedge, \vee, \top, \mathrm{I}, 0, 1)$ is a residuated lattice.

The *lattice of uncertainty values* $(V, \wedge, \vee, 0, 1, n_V, \top)$ will be a finite distributive lattice with two additional operations: an involution n_V and a t-norm $\top$ on V. The *lattice of preference values* $(U, \leq_U, 0, 1, n_U)$ will be a finite distributive lattice with involution. We will denote by $Pi(X)$ the set of *consistent possibility distributions on X over V*, i.e.

$$Pi(X) = \{\pi : X \to V \mid \bigvee_{x \in X} \pi(x) = 1\}.$$

The pointwise ordering on $Pi(X)$ is then defined as $\pi \leq \pi'$ iff $\forall x \in X$ $\pi(x) \leq_V \pi'(x)$, with $\leq_V$ the order induced by $\wedge$ in V. We will be interested in a subset of $Pi(X)$, the set of *normalised* possibility distributions, i.e. the set

$$Pi^*(X) = \{\pi \in Pi(X) \mid \exists x \text{ s.t. } \pi(x) = 1\}.$$

For the sake of simplicity, we shall use A for denoting both both a subset $A \subseteq X$ and the normalised possibility distribution on X such that $\pi(x) = 1$ if $x \in A$ and $\pi(x) = 0$ otherwise. Hence, we can consider X as included in $Pi^*(X)$.

3 Generalised Utility functions for $\vee$-Mixtures

As it has been mentioned, QU^- and QU^+ are "utility" functions in $Pi^*(X)$, in the sense that they preserve the preference ordering and the maximin combination of possibilistic mixture. Now, we analyse the conditions required to guarantee that the functions GQU^- and GQU^+ preserve a possibilistic mixture. Instead of applying maximin combination of possibility distributions, we consider other mixtures involving t-conorms and t-norms. For each t-norm $\top$ and conorm $\perp$ on V, we will be interested in $\perp$-$\top$ mixtures that combine two possibility distributions π_1 and π_2 into a new one, denoted $M_{\top,\perp}(\pi_1, \pi_2, \lambda, \mu)$, with $\lambda, \mu \in V$ and $\lambda \perp \mu = 1$, defined as

$$M_{\top,\perp}(\pi_1, \pi_2, \lambda, \mu)(x) = (\lambda \top \pi_1(x)) \perp (\mu \top \pi_2(x)).$$

We require these mixtures to satisfy reduction of lotteries, that is:

$$M_{\top,\perp}(M_{\top,\perp}(\pi_1, \pi_2, \lambda_1, \lambda_2), M_{\top,\perp}(\pi_1, \pi_2, \mu_1, \mu_2), \alpha, \beta)$$
$$= M_{\top,\perp}((\pi_1, \pi_2, (\alpha \top \lambda_1) \perp (\beta \top \mu_1), (\alpha \top \lambda_2) \perp (\beta \top \mu_2)).$$

Hence, we need to satisfy $(a \top c) \perp (b \top c) = c \top (a \perp b)$. Therefore, we have to restrict to $\vee - \top$ mixtures [1], moreover, we have to require $(V, \wedge, \vee, \top, \mathrm{I}, 0, 1)$ to be a residuated lattice, hence forth V will be referred as a residuated lattice. So, for each t-norm $\top$ on V, we may consider the so-called Possibilistic Mixture. In order to have a closed operation on $Pi^*(X)$, the mixture operation is restricted to $Pi^*(X)$ requiring the scalars to satisfy an additional condition $\lambda = 1$ or $\mu = 1$. As it is said, we will consider here other alternative for modelling implication, we define

$$(v \Rightarrow u) = n(v \top z)$$

with $n(z) = u$, $\top$ a t-norm on V and $n = n_U \circ h$, being n_U the involution in U, and $h : V \to U$ an order preserving function, such that $h(0) = 0, h(1) = 1$. Hence, given a preference function $u : X \to U$ that

assigns to each consequence of X a preference level of U, for a pessimistic behaviour we propose

$$GQU^-(\pi|u) = \bigwedge_{x\in X} n(\pi(x)\top\lambda_x).$$

being λ_x s.t. $n(\lambda_x) = u(x)$. To guarantee the correctness of the above definition of implication we require h to satisfy the following coherence condition w.r.t. $\top$,

$$h(\lambda) = h(\mu) \Rightarrow h(\alpha\top\lambda) = h(\alpha\top\mu) \quad \forall\alpha,\lambda,\mu \in V.$$

Notice that either when $\top = \wedge$ or when h is injective this condition is satisfied. If h is coherent w.r.t. $\top$, so is n. Observe that the implication proposed may be seen like a generalisation of a S-implication. Actually, when h is injective, h satisfies coherence, then

$$(v \Rightarrow u) = n(v\top z) = n(v)\perp_{n,\top} u,$$

being $\perp_{n,\top}$ the conorm in U defined as $(n(\lambda)\perp_{n,\top} n(\lambda')) = n(\lambda\top\lambda')$. That is, $(v \Rightarrow u)$ is a S-implication w.r.t. the conorm $\perp_{n,\top}$. Instead, for an optimistic behaviour we consider the t-norm as the conjunction, so we propose

$$GQU^+(\pi|u) = \bigvee_{x\in X} h(\pi(x)\top\mu_x)$$

being μ_x s.t. $u(x) = h(\mu_x)$. Observe that if h is join-preserving and as V is a residuated lattice with involution, then GQU^- and GQU^+ preserves the possibilistic mixture in the sense that the following expressions hold,

$$GQU^-(M_\top(\pi_1,\pi_2,\lambda,\mu)|u)(x) = n(\lambda\top\delta_1) \wedge n(\mu\top\delta_2),$$
$$GQU^+(M_\top(\pi_1,\pi_2,\lambda,\mu)|u)(x) = h(\lambda\top\gamma_1) \vee h(\mu\top\gamma_2),$$

with $n(\delta_j) = GQU^-(\pi_j|u)$, $h(\gamma_j) = GQU^+(\pi_j|u)$,

4 Representation of Generalised Qualitative Utilities

First, we propose a set of axioms to characterise pessimistic and optimistic qualitative utilities for normalised possibility distributions. Then, in order to may apply the model in contexts involving partially inconsistent belief states, we introduce utility functions for evaluating possibly non normalised distributions and their corresponding axiomatic characterisations.

Proposition 1 *Let $(Pi^*(X), \sqsubseteq)$, satisfying*

- *A1: $(Pi^*(X), \sqsubseteq)$ is a pre-lattice.*
- *A2(uncertainty aversion): if $\pi \leq \pi' \Rightarrow \pi \sqsupseteq \pi'$.*

Then

a) The maximal elements of $(Pi^*(X), \sqsubseteq)$ are equivalent.

b) The maximal elements of $(X, \sqsubseteq)$ are equivalent, and they are equivalent to the maximal ones of $(Pi^*(X), \sqsubseteq)$.

Axiomatic setting: Let $AXP_\top$ be the following set of axioms on $(Pi^*(X), \sqsubseteq, M_\top)$,

- A1: $(Pi^*(X), \sqsubseteq)$ is a pre-lattice.
- A2 (uncertainty aversion): if $\pi \leq \pi' \Rightarrow \pi \sqsupseteq \pi'$.
- A3 (independence): $\pi_1 \sim \pi_2 \Rightarrow M_\top(\pi_1,\pi,\lambda,\mu) \sim M_\top(\pi_2,\pi,\lambda,\mu)$.

Let $\overline{\pi}$ be a maximal element of $(Pi^*(X), \sqsubseteq)$, then $\pi_\lambda = M_\top(\overline{\pi}, X, 1, \lambda)$.

- A4: if $\pi_\lambda \sqsubseteq \pi'_\lambda \Rightarrow \pi_{n_V(\lambda)} \sqsupseteq \pi_{n_V(\lambda')}$,
- A5: if $\lambda <> \lambda' \Rightarrow \pi_\lambda \sqsubset\sqsupset \pi_\lambda$'.
- A6: $\forall\pi \in Pi^*(X), \exists\lambda \in V$ s.t. $\pi \sim M_\top(\overline{\pi}, X, 1, \lambda)$.

Lemma 2 *Let $(U, \leq_U, 0, 1, n_U)$ and $(V, \wedge, \vee, \top, I, 0, 1)$ be two lattices with involution, $h : V \to U$ an onto join-preserving mapping satisfying coherence w.r.t. $\top$, and $u : X \to U$. If $(GQU^-)^{-1}(1) \neq \emptyset$[1] and $(GQU^-)^{-1}(0) \neq \emptyset$, then*

- *there exists $x \in X$ s.t. $u(x) = 1, \bigwedge_{x\in X} u(x) = 0$*
- *GQU^- is onto*

Lemma 3 *Let $h : V \to U$ be an onto non decreasing function also satisfying that if $\lambda <> \lambda'$ then $h(\lambda) <>_U h(\lambda')$. Then, h is a preserving epimorphism.*

Now, let $\preceq_{GQU^-}$ be the preference ordering on $Pi^*(X)$ induced by GQU^-, i.e.

$$\pi \preceq_{GQU^-} \pi' \text{ iff } GQU^-(\pi|u) \leq_U GQU^-(\pi'|u).$$

Theorem 4 *A preference relation $(Pi^*(X), \sqsubseteq, M_\top)$ satisfies axioms $AXP_\top$ iff there exist*

- *a utility finite distributive lattice $(U, \wedge, \vee, n_U, 0, 1)$,*
- *a preference function $u : X \to U$, s.t. $u^{-1}(1) \neq \emptyset$ and $\bigwedge_{x\in X} u(x) = 0$,*
- *an onto join-morphism $h : V \to U$, i.e. $h(\lambda \vee \lambda') = h(\lambda) \vee h(\lambda')$, s.t. $h(0) = 1, h(1) = 1$, satisfying coherence w.r.t. $\top$, $n_U \circ h \circ n_V = h$, and also satisfying*

$$\text{if } \lambda <> \lambda' \text{ then } h(\lambda) <>_U h(\lambda'),$$

in such a way that it holds:

$$\pi' \sqsubseteq \pi \text{ iff } GQU^-(\pi|\, u) \leq_U GQU^-(\pi'|\, u).$$

In order to represent an optimistic preference criterion, we consider now the distribution π_λ defined as $\pi_\lambda = M_\top(X, \underline{\pi}, \lambda, 1)$, where $\underline{\pi}$ is minimal of $(Pi^*(X), \sqsubseteq)$, and we have to change the uncertainty aversion axiom A2 by an uncertainty-prone postulate

- A2$^+$: if $\pi \leq \pi'$ then $\pi \sqsubseteq \pi'$

and to modify the continuity axiom A6 into

- A6$^+$: $\forall\pi \in Pi^*(X)\exists\lambda \in V$ s.t. $\pi \sim M_\top(X, \underline{\pi}, \lambda,$

For an optimistic behaviour, we may consider $\preceq_{GQU^+}$ the preference ordering on $Pi^*(X)$ induced by GQU^+. The corresponding representation theorem is analogous to the pessimistic case and it is omitted.

Representation of Utilities for Non Normalised Distributions

Applying these models to non normalised distributions may conclude in unsatisfactory effects like that the

[1] For a simpler notation we will omit u when there is no confusion about it.

pessimistic utility may result higher than the optimistic utility. In order to avoid some problems involving non normalised distributions, we provide now the corresponding extension of our initial proposal. First, let us introduce the concepts of normalisation and height of a distribution. Let $\mathcal{H}$ be the height of a distribution, i.e. $\mathcal{H}(\pi) = \bigvee_{x \in X} \pi(x)$, for each distribution we consider the subset of consequences with maximal plausibility

$$X_{\mathcal{H}(\pi)} = \{x \in X | \ \forall y \in X \ \pi(y) \not> \pi(x)\}.$$

Define $\mathcal{N}(\pi)$ the normalisation of π, i.e. the normalised distribution

$$\mathcal{N}(\pi) = \begin{cases} 1 & if \quad x \in X_{\mathcal{H}(\pi)} \\ \pi(x) & otherwise \end{cases}$$

and consider the extension of the set of possibilistic lotteries to the set $Pi^{ex}(X)$ of non necessarily normalised distributions on V. Now, we propose the qualitative (or ordinal) utility functions on $Pi^{ex}(X)$,

$$\begin{aligned} \underline{GQU}^-(\pi|u) &= GQU^-(\mathcal{N}(\pi)|u) \wedge n(n_V(\mathcal{H}(\pi))) \\ \underline{GQU}^+(\pi|u) &= GQU^+(\mathcal{N}(\pi)|u) \vee h(n_V(\mathcal{H}(\pi))). \end{aligned}$$

Let $\sqsubseteq_{ex}$ be a preference relation in $Pi^{ex}(X)$. We will denote by $\sqsubseteq$ its restriction to $Pi^*(X)$, $\sim_{ex}$ and $\sim$ the corresponding indifference relations. In order to characterise the preference orderings induced by $\underline{GQU}^-$, we extend the axiom set $AXP_\top$, defined on $(Pi^*(X), \sqsubseteq, M_\top)$, with A7P:

- $\forall \pi \in \ Pi^{ex}(X) \pi \sim_{ex} M_\top(\mathcal{N}(\pi), X, 1, n_V(\mathcal{H}(\pi)))$.

The intuitive idea behind this axiom is that we make a non-normalised possibilistic lottery π indifferent to the corresponding normalised lottery $\mathcal{N}(\pi)$, provided that it is modified by a uniform uncertainty level corresponding to the inconsistency degree of π.
We say that $\sqsubseteq_{ex}$ on $Pi^{ex}(X)$ satisfies axiom set $\underline{AXP}_\top = AXP_\top \cup \{A7P\}$ iff $\sqsubseteq$ satisfies $AXP_\top$ and $\underline{\sqsubseteq_{ex}}$ also satisfies A7P.

Theorem 5 *A preference relation $\underline{\sqsubseteq_{ex}}$ on $Pi^{ex}(X)$ satisfies axiom set $\underline{AXP}_\top$ iff there exist*

- *a utility finite distributive lattice $(U, \wedge, \vee, n_U, 0, 1)$ with involution n_U,*
- *a preference function $u : X \to U$, s.t. $u^{-1}(1) \neq \emptyset$ and $\bigwedge_{x \in X} u(x) = 0$,*
- *a preserving epimorphism $h : V \to U$ s.t. $h(0) = 0$, $h(1) = 1, n_U \circ h \circ n_V = h$, satisfying coherence w.r.t. $\top$, and if $\lambda <> \lambda'$ then $h(\lambda) <>_U h(\lambda')$,*

in such a way that it holds:

$$\pi' \sqsubseteq_{ex} \pi \ iff \ \underline{GQU}^-(\pi') \leq_U \underline{GQU}^-(\pi).$$

In a similar way, we characterise the "optimistic" utility on $Pi^{ex}(X)$, using the set $AXP_\top^+$ plus A7P.

Remark 1 . *Instead of using the involution n_V in the definition of $\underline{GQU}^-$ and $\underline{GQU}^+$, one could simply use a more general function $F : V \to V$ s.t. $F(1) = 0$,*

$$\begin{aligned} \underline{GQU_F}^-(\pi) &= GQU^-(\mathcal{N}(\pi)) \wedge n(F(\mathcal{H}(\pi))) \\ \underline{GQU_F}^+(\pi) &= GQU^+(\mathcal{N}(\pi)) \vee h(F(\mathcal{H}(\pi))). \end{aligned}$$

In that case, given such a function F, it is not difficult to show that Theorem 5 is still valid provided that we replace axiom A7P by an analogous one:

- $\forall \pi \in \ Pi^{ex}(X) \ \pi \sim_{ex} M_\top(\mathcal{N}(\pi), X, 1, F(\mathcal{H}(\pi)))$

5 Conclusions

We have been concerned with representational aspects of preference relations in the framework of a possibilistic (qualitative / ordinal) decision model under uncertainty. The model measures degrees of uncertainty (similarity) and preferences on partially ordered sets (finite lattices). Here, we assumed the availability of t-norms operators as well as the usual $\wedge, \vee$ ones. This allowed us to apply non purely ordinal expressions in the utility functions involved. Then, we extended the model to cope with decision problems where the belief state may be partially inconsistent, like those involved in case-based decision problem. This enables the model to rank, possibly non-normalised, possibility distributions on the set of decision consequences.

Acknowledgements

This work has been partially supported by the CICYT project SMASH(TIC96-1138-C04-01) and A. Zapico acknowledges partial support by the Univ. Nac. de Rio Cuarto (Argentina).

References

[1] G. De Cooman and E. Kerre. Order norms on bounded partially ordered sets. *J. of Fuzzy Math*, 2:281–310, 1993.

[2] B. Davey and H. Priestley. *Introduction to Lattices and Order*. Cambridge Univ. Press, 1990.

[3] H. Prade, D. Dubois, Ll. Godo and A. Zapico. Making decision in a qualitative setting: from decision under uncertainty to case-based decision. In *6th Int. Conf. on Principles of Knowl. Representation and Reasoning(KR'98)*, pp. 594 - 605, 1998.

[4] D. Dubois and H. Prade. Possibility theory as a basis for qualitative decision theory. In *14th Int. Joint Conf. on Artif. Intell. (IJCAI'95)*, pp. 1924–1930,1995.

[5] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princenton Univ. Press, 1944.

[6] A. Zapico. Axiomatic foundations for qualitative/ordinal decisions with partial preferences. In *16th Int. Joint Conf. on Artif. Intell.(IJCAI'99)*.

[7] A. Zapico and Ll. Godo. On the possibilistic-based decision model: preferences under partially inconsistent belief states. In *Workshop on Decision theory meets AI: qualitative and quantitative approaches.ECAI'98*, pp. 99–109, 1998.

CLASSIFYING QUALITATIVE INFORMATION USING CENTROID BASED METHODS

Vicenç Torra, Lluís Godo

Institut d'Investigació en Intel·ligència Artificial - CSIC

Campus UAB s/n, 08193 Bellaterra (Catalunya, Spain)

{vtorra, godo}@iiia.csic.es

Abstract

In this paper we consider the problem of classification of qualitative data. We present an approach to overcome several difficulties which show SAHN (sequential, agglomerative, hierarchic, nonoverlapping) clustering methods to classify qualitative data when using centroids to compute similarities between pairs of classes. The approach is based on a recently proposed class of qualitative weighted average functions.

Keywords: qualitative aggregation, centroid, classification.

1 INTRODUCTION

Classification is one of the main tasks which is present in many AI systems ranging from knowledge acquisition tools to knowledge based systems verification. A classification process usually consists [Everitt, 1977] of a three step process: (1) domain description, (2) similarity construction and (3) classification construction. There are a lot of available classification methods, even if the data is imprecisely known and modelled as fuzzy sets, e.g. [Bezdek, 1981].

However, almost all the methods only deal with numerical data [Gordon, 1996], and if the available data to be classified is only of qualitative nature, a numerical or fuzzy sets interpretation of the data is required.

Among the algorithms to build classification trees we can distinguish the set of methods known as SAHN [Everitt, 1977; Valls et al., 1997] i.e., sequential, agglomerative, hierarchic, and nonoverlapping clustering methods. Given a set of objects $O = \{o_1, ..., o_m\}$ to classify in a tree of classes C, these methods follow the structure given below:

1. For each object o_i, define a class that consists only of that object
2. Computation of the similarities among all pairs of classes
3. While there exists more than a single class in C
   - 3.1. Selection of the classes K that should integrate the new class
   - 3.2. Creation of a class C_K with the classes in K
   - 3.3. Removal of the classes in K
   - 3.4. Calculation of the similarity between C_K and the ones in C
4. End while

This can be expressed more conscisely as follows:

1. For all $o_i \in O$ loop $C = \cup \{o_i\}$ end loop
2. S = Similarity (C)
3. while $|C| > 1$
   - 3.1. C_K = Selection (S, C), $C_K \subseteq O$
   - 3.2. $C = C \cup \{C_K\}$
   - 3.3. $C = C - \{c \in C \mid c \subset C_K, c \neq C_K\}$
   - 3.4. S = NewSimilarity(S, C, C_K)
4. end while

In this general approach, two aspects are left to the user: the selection of the aggregation criterion (how to select the objects that form the new class), and the classification method (how to calculate new similarities from old ones). Several alternatives exist for both criteria. Among them, we can distinguish, in relation to the former, the selection of all those elements that are related with a large similarity. In relation to the latter criteria, one of the existing methods is the Centroid cluster analysis. In this case, to compute the similarity between two classes, first it is computed a representative of the class (the so-called *centroid*), and then a similarity function is applied to the centroids. See [Aguilar, 1995, and Aguilar et al., 1990] for an interesting and similar approach to classification of qualitative data based on adequation degrees of objects to clusters.

As a matter of notation, in this paper we consider that the set of relevant attributes to be used in the classification is $A = \{A_1, ..., A_n\}$, with domains $Dom(A_i)$ consisting of a finite set of linearly ordered (linguistic) labels. We use $Val(o_i, A_j) \in Dom(A_j)$ to denote the evaluation of the object o_i in relation to the attribute A_j.

In general, using the centroid method to compute the similarity between two classes C_R and C_P built from objects $R = \{r_1, ..., r_r\}$ and $P = \{p_1, ..., p_p\}$ we proceed as follows:

1. Computation of the centroids of R and P

$$\text{Centroid(R)} = (\mathbb{C}_{A1}(\text{Val}(r_i, A_1)_{i \in \{1,\ldots,r\}}), \ldots, \mathbb{C}_{An}(\text{Val}(r_i, A_n)_{i \in \{1,\ldots,r\}}))$$

$$\text{Centroid(P)} = (\mathbb{C}_{A1}(\text{Val}(p_i, A_1)_{i \in \{1,\ldots,p\}}), \ldots, \mathbb{C}_{An}(\text{Val}(p_i, A_n)_{i \in \{1,\ldots,p\}}))$$

where $\mathbb{C}_{Ai}$ is an aggregation operation (usually the arithmetic mean) defined on the domain of the attribute A_i. Note that this aggregation operator has to be defined in such a way that can combine an arbitrary number of linguistic labels. This is due to the fact that not all classes have always the same number of objects.

To simplify the notation in the rest of the work, we will use Val(P, A_k) and Val(R, A_k) to denote:

$$\text{Val(P, } A_k) = \mathbb{C}_{Ak}(\text{Val}(p_i, A_k)_{i \in \{1,\ldots,p\}})$$
$$\text{Val(R, } A_k) = \mathbb{C}_{Ak}(\text{Val}(r_i, A_k)_{i \in \{1,\ldots,r\}})$$

2. Computation of the multi-dimensional similarity between the two centroids

$$\text{Multi-dimensional-similarity(R, P)} = (S_{A1}(\text{Val(R, } A_1), \text{Val(P, } A_1)), \ldots, S_{An}(\text{Val(R, } A_n), \text{Val(P, } A_n)))$$

3. Reduction of the multi-dimensional similarity to a one-dimension similarity.

$$\text{S(R, P)} = \text{Agg(Multi-dim-similarity(R, P))}$$

where Agg is an aggregation operator.

In the numerical case, the computation of all these values is straightforward. In that case, $\mathbb{C}_{Ai}$ can be any numerical aggregation operator (e.g. the arithmetic mean), S_{Ai} is any similarity function (e.g. a dual of a normalized Euclidean distance) and the aggregation procedure in step 3 can be again any numerical aggregation operator, since all the values S_{Ai}(Val(R, A_i), Val(S, A_i)) to aggregate are numerical.

2 QUALITATIVE CLASSIFICATION

In the case attributes take values in an ordinal scale, without any underlying numerical or fuzzy set based representation, difficulties arise at the three stages considered above. Let us consider them in more detail:

1. *Computation of the centroid.* In the qualitative setting, we would need aggregation procedures $\mathbb{C}_{Ai}$ able to combine qualitative terms instead of numerical ones. Besides of that, procedures $\mathbb{C}_{Ai}$ depend on the domain of the attribute A_i. Therefore we would need to define as much operators as attributes we have.

2. *Computation of the multi-dimensional similarity.* As in the previous case, it is required a similarity function for each existing domain.

3. *Reduction to a one-dimension similarity.* The way this reduction is performed depends on how functions S_{Ai} are defined. Note that, in general, given two linguistic terms *a* and *b* in the domain of A_i, the function S_{Ai} will compute its similarity in a certain domain D. However, it is a common assumption that the domain D coincides with the domain of A_i. In this case, the reduction function has to combine n values in n different domains.

Below we briefly describe our approach to face each of these problems.

2.1 CENTROID COMPUTATION

To define the centroid of a class we use, for each attribute, a qualitative aggregation operator based on the qualitative weighted mean defined in [Godo and Torra, 1998, 1999].

This operator performs a weighted aggregation of values u_i belonging to a certain ordinal scale $U=\{0=u_0 < \ldots < 1=u_n\}$ and weighted by natural numbers. We will denote by $\mathbb{N}$ the set of natural numbers with the usual addition (+) and substraction (-). To overcome the difficulty of having two different domains, the approach consists of building a new unified domain $U_{\mathbb{N}} = \mathbb{N} \times (U-\{1\})$ where suitable addition, product and division like operators needed for the aggregation are defined.

As for the addition operator in $U_{\mathbb{N}}$, we need, first of all, an operation in U accounting for the notion of accumulation. It is well known on fuzzy set theory that t-conorms are the binary operations which are closest to this type of operations, enjoying properties like commutativity and associativity, specially needed in the qualitative setting.

From now on $\oplus$: $U \times U \to U$ will denote a finite t-conorm on the scale (U, <), that is, a non-decreasing commutative and associative operation fulfilling these boundary conditions: for all $u \in U$, $u \oplus 0 = u$ and $u \oplus 1 = 1$. Moreover, if we denote by n_U the order-reversing involution on U, one can define the t-norm $\otimes$: $U \times U \to U$ which is the dual operation of $\oplus$ with respect to n_U as:

$$u \otimes v = n_U(n_U(u) \oplus n_U(v)).$$

Once operations $\oplus$ and $\otimes$ are defined, we proceed to embed the algebraic structure $(U, <, \oplus, \otimes)$ into $U_{\mathbb{N}}$. The transformation T: $U \to U_{\mathbb{N}}$ that maps values of U into values of $U_{\mathbb{N}}$ is as follows:

$$T(u) = \begin{cases} (0,u), & \text{if } u<1 \\ (1,0), & \text{if } u=1 \end{cases}$$

The orderings in U and in $\mathbb{N}$ induce the following lexicographic ordering in $U_{\mathbb{N}}$:

$$(n, u) \leq (m, v) \text{ if either}$$

$$n < m \text{ or } (n = m \text{ and } u \le v).$$

Next, a corresponding binary addition-like operation in the extended domain

$$\underline{\oplus}: U_{\mathbb{N}} \times U_{\mathbb{N}} \to U_{\mathbb{N}},$$

extending the $\oplus$ operation in U, is defined in such a way that keeps track, by means of the t-norm, of how much the accumulation exceeds from 1.

Definition 1. $(n,x) \underline{\oplus} (m,y) =$

$$\begin{cases} (n+m,\ x \oplus y), & \text{if } x \oplus y < 1 \\ (n+m+1,\ x \otimes y), & \text{if } x \oplus y = 1 \end{cases}$$

In [Godo and Torra, 1999] necessary conditions for the t-conorm $\oplus$ are given so that the accumulation operation $\underline{\oplus}$ be associative, and thus, it is enough to have it defined as a binary operation.

The weighting of elements of U by natural numbers is done by means of a product-like operation

$$\bullet : \mathbb{N} \times U_{\mathbb{N}} \to U_{\mathbb{N}},$$

defined as an iterative procedure of the addition operation $\underline{\oplus}$ in the natural way.

Definition 2. For each n, $m \in \mathbb{N}$ and $x \in U-\{1\}$, we define:

$$n\bullet(m, x) = (m, x) \underline{\oplus} \overset{n)}{..} \underline{\oplus} (m, x).$$

Notice that $(n_1+n_2)\bullet(m, x) = n_1\bullet(m, x) \underline{\oplus} n_2\bullet(m, x)$ and that $n\bullet[(m_1, x) \underline{\oplus} (m_2, y)] = n\bullet(m_1, x) \underline{\oplus} n\bullet(m_2, y)$.

Finally, to complete the modelling of the averaging of the values $(u_1, ..., u_m)$ with weights $(n_1, ..., n_m)$, a division-like operation $\varnothing$ between the extended value resulting from the addition $[n_1\bullet T(u_1)] \underline{\oplus} ... \underline{\oplus} [n_m\bullet T(u_m)]$ and the sum of weights $n_1 + ... + n_m$ is defined.

Definition 3. For each n, $m \in \mathbb{N}$ and $x \in U-\{1\}$, we define $(n, x) \varnothing m = ([n/m], y)$ being $[n / m] = \max\{p \in \mathbb{N} \mid p \cdot m \le n\}$ (integer division) and $y = \max\{z \in U-\{1\} \mid m\bullet(0, z) \le (r, x)\}$, where $r = n - [n / m]\cdot m$.

Finally, since the aggregation should be in U, we use T^{-1} to yield the final result.

Definition 4. Given a t-conorm $\oplus$ in a finite scale U inducing operations ($\underline{\oplus}$, $\bullet$, $\varnothing$) as defined above, the **qualitative weighted mean** of a vector $\mathbf{u} = (u_1, ..., u_m)$ of values of U with respect to a vector $\alpha = (n_1, ..., n_m)$ of weights of $\mathbb{N}$ is defined as

$$QWM_{U,\oplus}(\alpha,\mathbf{u}) = T^{-1}(([n_1\bullet T(u_1)] \underline{\oplus} ... \underline{\oplus} [n_m\bullet T(u_m)]) \varnothing \Sigma_{i=1,m}\, n_i)$$

In our case, to compute the centroid we assume that $n_i=1$ for all i. This means that all objects in the class are equally represented in the centroid.

This definition depends on the domain U and also on the t-conorm $\oplus$. Therefore to use it to compute the centroid we need to settle both of them. Thus we have:

$$\mathbb{C}_{A_i}(\mathbf{u}) = QWM_{DOM(A_i),\oplus}(\mathbf{1}, \mathbf{u})$$

where $\mathbf{1} = (1, ..., 1)$.

2.2 COMPUTATION OF THE MULTI-DIMENSIONAL SIMILARITY

In this case, we have defined a similarity function for each attribute as a many-valued equivalence connective. We have based the similarity on the same t-conorm used to define the aggregation operator to compute the centroid.

$$S_{A_i}(x, y) = \min(\max\{z: (x \otimes z) \le y\}, \max\{z: (y \otimes z) \le x\})$$

In this function, $\otimes$ is the dual t-norm of $\oplus$. Notice that $S_{A_i}(x, x)$ takes the maximum value of the domain. Thus, the function depends on both the domain of the attribute and the t-conorm $\oplus$. Therefore, in general we have a function for each attribute. Note also that $S_{A_i}(x,y)$ is in the domain of A_i.

2.3 REDUCTION TO A ONE-DIMENSIONAL SIMILARITY

To reduce to a one-dimension similarity, we need to introduce a new set of labels D_c as a common domain for all attributes. To this end we take D_c large enough to allow each domain $Dom(A_i)$ to be embedded into D_c by means of an onto order-preserving mapping f_{A_i}: $Dom(A_i) \to I(D_c)$, sending each domain value to an interval of values of D_c, and such that $\cup\{f_{A_i}(a) \mid a \in Dom(A_i)\} = D_c$. Here $I(D_c)$ denotes the set of intervals of D_c.

Let us assume that the multi-dimensional-similarity between two centroids R and P is given by the vector $(a_1, ..., a_n)$, where

$$a_i = S_{A_i}(Val(R, A_i), Val(P, A_i)) \in Dom(A_i)$$

Then, the one-dimensional similarity between R and P is defined as

$$S(R, P) = Agg(a_1, ..., a_n) = Agg'(f_{A_1}(a_1), ..., f_{A_n}(a_n))$$

A particular definition of the Agg' function could be just to let

$$Agg'(f_{A_1}(a_1), ..., f_{A_n}(a_n)) = Min\{S'(f_{A_i}(a_i), f_{A_j}(a_j)) \mid i \ne j\}$$

where $S'([I^i, I^s], [J^s, J^s]) = \min(S^*(I^i, J^i), S^*(I^s, J^s))$, being S^* a similarity defined (in the same way as in Section 2.2) on D_c.

3 EXAMPLE

We have applied the method developed to a small example. The example consists of only 10 objects (cars), denoted by {c1, c2, .., c10} and extracted from a large file of 1728 objects [Murphy and Aha, 1994], that are evaluated using 6 qualitative attributes: buying price, price of maintenace, number of doors, capacity in terms of persons to carry, size of luggage boot, and estimated safety of the car. The domain of these attributes is given in Table I and the evaluation of the 10 objects is given in Table II.

Attribute	Domain
buying	low med high vhigh
maint	low med high vhigh
doors	two three four 5more
persons	two four more
lug_boot	small med big
safety	low med high

Table I. Attribute domains

	buying	maint	doors	pers.	lug-b.	safety
c1	vhigh	vhigh	two	two	small	low
c2	vhigh	vhigh	three	two	small	low
c3	vhigh	vhigh	three	two	small	med
c4	vhigh	vhigh	four	two	small	med
c5	vhigh	vhigh	5more	two	small	low
c6	high	low	5more	more	small	low
c7	high	low	5more	more	small	med
c8	high	low	four	more	small	med
c9	high	med	four	four	small	high
c10	high	med	three	four	small	high

Table II. Objects to classify

We have defined the aggregation operator in each domain based on the corresponding Lukasiewicz t-conorm ($a_i \oplus a_j = a_{\min(i+j,n)}$, where n is the cardinality of the domain).

The common domain D_c where all the attribute domains in Table I have been mapped has been taken as a 12 element scale, in such a way that each element of buying, maint and doors domains are mapped to a three element intervals, while each element of the remaining domains are mapped to a 4 element interval.

The classification we have obtained for these objects is given in Figure 1.

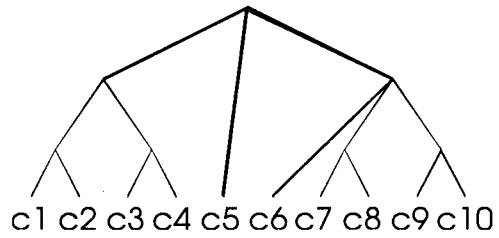


Figure 1. Classification of data in Table I

4 CONCLUSIONS AND FUTURE WORK

We have described how to apply a class of qualitative aggregation procedures in a classifier system based on centroids. The methodology developed permits to operate all the elements while remaining in the qualitative setting without translating them into quantitative one.

As a future work it is needed a deeper study of the process of reduction from a multi-dimensional similarity to a one dimensional similarity. With the approach used here we need to define a unified framework and a transformation of each value in each domain into this unified framework. This assumption needs, however, too much information to be supplied by domain experts. An alternative approach with no so much knowledge would be better for real applications.

Acknowledgements

The authors acknowledge partial support from the CICYT project SMASH, TIC96-1138-C04-01/04.

References

Aguilar-Martin, J., (1995), *Representation inductive des connaissances a partir des donnees,* Report de Recherche, LAAS.

Aguilar-Martin, J., Martín, M., Piera, N., (1990), Conceptual Connectivity Analysis by Means of Fuzzy Partitions, *Proc. of the IPMU 90*, pp. 250-252, Paris, France.

Bezdek J.C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, NY.

Cignoli R., D'Ottaviano I., Mundici D. (1995), Algebras das logicas de Lukasiewicz, *Coleçao Centro de Logica, Epistemologia e Historia da Ciencia*, Unicam (Brasil) Volumen 12 (1995).

Everitt B., (1977) *Cluster Analysis*, Heinemann Educational Books Ltd.

Godo L., Torra, V., (1998), On qualitative Weighted Means, *Actas del VIII Congreso español de tecnologias y lógica fuzzy*, (ISBN 84-95075-10-5), 185-192, Pamplona.

Godo L., Torra V., (1999) On aggregation operators for ordinal qualitative information, submitted.

Gordon A. D., (1996) Hierarchical Classification. In *Clustering and Classification*, P. Arabie, L.J. Hibert and G. De Soete (eds.), World Scientific Publishing, River Edge NJ, pp. 65-121.

Murphy P.M., Aha D.W. (1994) UCI Repository machine learning databases, Univeristy of California, Irvine, CA. http://www.ics.uci.edu/~mlearn/MLRepository.html,

Valls, A., Riaño, D., Torra, V., (1997), Sedàs: A semantic based general classifier system, *Mathware and Soft Computing*, 4, 267-279.

TRIANGULAR NORMS ON THE REAL UNIT SQUARE

Bernard De Baets
Department of Applied Mathematics
and Computer Science
University of Ghent
Krijgslaan 281 (S9)
B-9000 Gent, Belgium
E-mail: Bernard.DeBaets@rug.ac.be

Radko Mesiar
Department of Mathematics
Slovak Technical University
Radlinského 11
SK-81368 Bratislava, Slovakia
E-mail: mesiar@sol.svf.stuba.sk

ÚTIA
Academy of Sciences of the Czech Republic
P.O. Box 18, 18208 Prague, Czech Republic

Summary

T-norms with vector arguments are introduced and investigated. Direct products of t-norms on the unit interval are characterized. T-norms without zero divisors, pseudo-Archimedean t-norms and cancellative t-norms are discussed. Some open problems are stated.

Keywords: cancellation law; direct product; pseudo-Archimedean property; t-norm; unit square.

1 INTRODUCTION

T-norms were introduced by Schweizer and Sklar (see e.g. [16]) in the framework of probabilistic metric spaces and are based on a notion used by Menger in order to extend the triangle inequality in the definition of metric spaces towards probabilistic metric spaces [13]. They have been used extensively for defining the intersection of fuzzy sets, and for modelling the logical 'and' in fuzzy logic. Continuous t-norms on the real unit interval have been completely characterized as ordinal sums of continuous Archimedean t-norms [11]. The structure of non-continuous t-norms is not known yet [10].

Due to the close connection between fuzzy set theory and order theory (see e.g. [8]), several authors have studied or worked with t-norms on bounded posets (see e.g. [3, 4, 5, 6, 9, 12, 14, 15] and many others). In this paper, we continue our study of t-norms on product lattices and focus on the particular case of the (real) unit hypercube $[0,1]^n$, i.e. we consider t-norms acting on vectors. From a practical point of view, the unit hypercube is the ultimate working environment: at the same time it allows for incomparability and also offers the possibility to draw back on the underlying real line.

2 T-NORMS ON $[0,1]^2$

T-norms can be introduced without any problem in a partially ordered framework. A t-norm T on a bounded poset $\mathbb{P} = (P, \leq, 0, 1)$ is defined as a commutative and associative increasing binary operation on P with neutral element 1. The partial order relation of $\mathbb{P}$ can be extended in a natural way to a partial order relation on the class of t-norms on it. For two t-norms T_1 and T_2 on $\mathbb{P}$, we define:

$$T_1 \leq T_2 \Leftrightarrow (\forall (x,y) \in P^2)(T_1(x,y) \leq T_2(x,y)).$$

We are now interested in the particular case of the (real) unit hypercube $([0,1]^n, \leq)$. For the sake of simplicity, we will write all results for $n = 2$, i.e. we will consider the (real) unit square $([0,1]^2, \leq)$ equipped with the product ordering $\leq$ defined by

$$(x_1, y_1) \leq (x_2, y_2) \Leftrightarrow x_1 \leq x_2 \wedge y_1 \leq y_2.$$

The unit square is a bounded lattice with smallest element $(0,0)$, greatest element $(1,1)$ and meet operation $\frown$ and join operation $\smile$ defined by

$$(x_1, y_1) \frown (x_2, y_2) = (\min(x_1, x_2), \min(y_1, y_2))$$
$$(x_1, y_1) \smile (x_2, y_2) = (\max(x_1, x_2), \max(y_1, y_2)).$$

The smallest t-norm T_W on $[0,1]^2$ is given by

$$T_W((x_1,y_1),(x_2,y_2)) = \begin{cases} (x_1,y_1) & \text{, if } (x_2,y_2) = (1,1) \\ (x_2,y_2) & \text{, if } (x_1,y_1) = (1,1) \\ (0,0) & \text{, elsewhere} \end{cases}$$

and the greatest t-norm T_M is given by the meet operation $\frown$, i.e. for any t-norm T on $[0,1]^2$ it holds that $T_W \leq T \leq T_M$.

3 DIRECT PRODUCTS OF T-NORMS

One easily verifies that for two t-norms T_1 and T_2 on $[0,1]$, the *direct product* $T_1 \times T_2$ defined by

$$T_1 \times T_2((x_1,y_1),(x_2,y_2)) = (T_1(x_1,x_2),T_2(y_1,y_2))$$

is a t-norm on $[0,1]^2$. Clearly, the smallest t-norm T_W is not a direct product of t-norms on $[0,1]$. Recall the following characterization of direct products.

Theorem 3.1 [4] *Consider a t-norm T on $[0,1]^2$, then the following statements are equivalent:*

(i) *T is the direct product of two t-norms T_1 and T_2 on $[0,1]$;*

(ii) *the partial mappings of T are meet-morphisms;*

(iii) *the partial mappings of T are join-morphisms.*

The foregoing theorem is quite important. It implies that in a residual framework, i.e. when the partial mappings of the t-norm are assumed to be supremum-morphisms, only direct products can be used.

4 BASIC PROPERTIES OF T-NORMS

In [4], we have introduced the sets of idempotent elements, zero divisors and nilpotent elements associated with an arbitrary t-norm T on a bounded poset $\mathbb{P} = (P, \leq, 0, 1)$:

(i) An element $x \in P$ is called an *idempotent element* of T if $T(x,x) = x$. The set of idempotent elements of T is denoted by $I(T)$; 0 and 1 are called trivial idempotent elements.

(ii) An element $x \in P$ is called a *zero divisor* of T if there exists $y \in P$ such that $\mathcal{L}(\{x,y\}) \neq \{0\}$ and $T(x,y) = 0$. The set of zero divisors of T is denoted by $Z(T)$; if $Z(T) = \emptyset$ then T is called a t-norm without zero divisors.

(iii) An element $x \in P \setminus \{0\}$ is called a *nilpotent element* of T if there exists $n \in \mathbb{N}$ such that $x^{(n)_T} = 0$. The set of nilpotent elements of T is denoted by $N(T)$.

Note that the notation $\mathcal{L}(A)$ stands for the set of lower bounds of the subset A of P. The notation $x^{(n)_T}$ stands for $T(x,\ldots,x)$ (n times argument x). By convention, we have that $x^{(1)_T} = x$.

In the case of the unit square, it holds that (x_1,y_1) is a zero divisor of a t-norm T on $[0,1]^2$ if and only if there exists (x_2,y_2) such that $(x_1,y_1) \frown (x_2,y_2) \neq (0,0)$ and $T((x_1,y_1),(x_2,y_2)) = (0,0)$. Note that since $T \leq T_M$, $(x_1,y_1) \frown (x_2,y_2) = (0,0)$ always implies that $T((x_1,y_1),(x_2,y_2)) = (0,0)$. Hence, such elements (x_2,y_2) cannot influence the decision whether (x_1,y_1) is to be considered as a zero divisor.

Consider two t-norms T_1 and T_2 on $[0,1]$ and their direct product $T_1 \times T_2$ on $[0,1]^2$. Then the following equalities hold:

$$I(T_1 \times T_2) = I(T_1) \times I(T_2)$$
$$N(T_1 \times T_2) = ((N(T_1) \cup \{0\}) \times (N(T_2) \cup \{0\})) \setminus \{(0,0)\}$$
$$Z(T_1 \times T_2) = (Z(T_1) \times [0,1]) \cup ([0,1] \times Z(T_2)).$$

In particular, the first equality above implies that the Cartesian product of the sets of trivial idempotent elements is included in the set of idempotent elements of any direct product of t-norms. Hence, the direct product $T_1 \times T_2$ not only has $(0,0)$ and $(1,1)$ as idempotent elements, but also $(0,1)$ and $(1,0)$. The third equality implies that the direct product $T_1 \times T_2$ is a t-norm without zero divisors if and only if both T_1 and T_2 are t-norms without zero divisors.

5 PSEUDO-ARCHIMEDEAN T-NORMS

The definition of the Archimedean property of binary operators on a bounded lattice (see e.g. [1]) can be applied to t-norms on a bounded poset as well. A t-norm T on a bounded poset $\mathbb{P} = (P, \leq, 0, 1)$ is called *Archimedean* if for any $(x,y) \in P^2$ it holds that

$$(\forall n \in \mathbb{N})(x^{(n)_T} \geq y) \Rightarrow (x = 1 \vee y = 0).$$

Clearly, a t-norm T on $[0,1]$ is Archimedean if and only if

$$(\forall (x,y) \in]0,1[^2)(\exists n \in \mathbb{N})(x^{(n)_T} < y).$$

One easily verifies that a t-norm T on a bounded poset $\mathbb{P} = (P, \leq, 0, 1)$ is Archimedean if and only if

$$(\forall x \in P \setminus \{1\})(\mathcal{L}(\{x^{(n)_T} \mid n \in \mathbb{N}\}) = \{0\}).$$

Of course, we would like to see the Archimedean property preserved under direct products. Consider the direct product $T = T_1 \times T_2$ of two t-norms T_1 and T_2 on $[0,1]$. As mentioned above, $(0,1)$ and $(1,0)$ are both idempotent elements of T. Hence, for any $(x,1) \in [0,1]^2$ the monotonicity of T implies that $(0,1) \in \mathcal{L}(\{(x,1)^{(n)_T} \mid n \in \mathbb{N}\})$. Similarly, for any $(1,y) \in [0,1]^2$ it holds that $(1,0) \in \mathcal{L}(\{(1,y)^{(n)_T} \mid n \in \mathbb{N}\})$. Therefore, elements of the type $(x,1)$ or $(1,y)$ cannot be nilpotent elements and T cannot be Archimedean. In order to avoid this undesirable situation, we have modified the definition of the

Archimedean property of t-norms on $[0,1]^2$. It is clear that we have to exclude from our considerations the elements greater than or equal to one of the idempotent elements $(0,1)$ and $(1,0)$. More formally, this is the set

$$U = (\{1\} \times [0,1]) \cup ([0,1] \times \{1\}).$$

Definition 5.1 *A t-norm T on $[0,1]^2$ is called pseudo-Archimedean if for any $(x,y) \in [0,1]^2 \times [0,1]^2$ it holds that*

$$(\forall n \in \mathbb{N})(x^{(n)_T} \geq y) \Rightarrow (x \in U \vee y = (0,0)).$$

Proposition 5.1 *Consider two t-norms T_1 and T_2 on $[0,1]$, then the direct product $T_1 \times T_2$ is pseudo-Archimedean if and only if T_1 and T_2 are Archimedean.*

6 THE CANCELLATION LAW

In a similar way, the cancellation law can be modified in order to guarantee that the direct product of cancellative t-norms is again cancellative. Recall that a t-norm T on $[0,1]$ is called *cancellative* if for any $x \neq 0$ and any $(y,z) \in [0,1]^2$ the equality $T(x,y) = T(x,z)$ implies $y = z$. In this case, a t-norm T is cancellative if and only if its partial mappings $T(x,\cdot)$, $x \neq 0$, are strictly increasing.

As mentioned above, it holds for any t-norm T on $[0,1]^2$ that $T((x_1,y_1),(x_2,y_2)) = (0,0)$ whenever $(x_1,y_1) \frown (x_2,y_2) = (0,0)$. This implies that we have to exclude from our considerations those elements (x_1,y_1) for which there exists $(x_2,y_2) \neq (0,0)$ such that $(x_1,y_1) \frown (x_2,y_2) = (0,0)$. More formally, this is the set

$$Z = (\{0\} \times [0,1]) \cup ([0,1] \times \{0\}).$$

Definition 6.1 *A t-norm T on $[0,1]^2$ is called cancellative if for any $x \in [0,1]^2 \setminus Z =]0,1]^2$ it holds that*

$$(\forall (y,z) \in [0,1]^2 \times [0,1]^2)(T(x,y) = T(x,z) \Rightarrow y = z).$$

This definition implies that the partial mappings $T(x,\cdot)$, $x \notin Z$, of a cancellative t-norm T are strictly increasing.

Proposition 6.1 *Consider two t-norms T_1 and T_2 on $[0,1]$, then the direct product $T_1 \times T_2$ is cancellative if and only if T_1 and T_2 are cancellative.*

7 TRANSFORMATIONS OF T-NORMS

One easily verifies that any order-preserving permutation (automorphism) φ of $[0,1]^2$, i.e. a permutation of $[0,1]^2$ that satisfies

$$(\forall (x,y) \in [0,1]^2 \times [0,1]^2)(x \leq y \Leftrightarrow \varphi(x) \leq \varphi(y)),$$

transforms any t-norm T on $[0,1]^2$ into a t-norm T_φ on $[0,1]^2$ defined by

$$T_\varphi(x,y) = \varphi^{-1}(T(\varphi(x),\varphi(y))).$$

Remarkably, automorphisms of the unit square are built up from automorphisms of the real unit interval, as follows: a $[0,1]^2 \to [0,1]^2$ mapping φ is an automorphism of $[0,1]^2$ if and only if there exist two automorphisms φ_1 and φ_2 of $[0,1]$ such that

$$(\forall (x,y) \in [0,1]^2)(\varphi(x,y) = (\varphi_1(x),\varphi_2(y)))$$

or

$$(\forall (x,y) \in [0,1]^2)(\varphi(x,y) = (\varphi_1(y),\varphi_2(x))).$$

One then easily verifies that for any automorphism φ of $[0,1]^2$ the transformation T_φ of the direct product $T = T_1 \times T_2$ of two t-norms T_1 and T_2 on $[0,1]$ is again a direct product of two t-norms on $[0,1]$.

8 CONTINUITY ISSUES

We have already recalled that continuous t-norms on $[0,1]$ are completely characterized as ordinal sums of continuous Archimedean t-norms. It is important to note that the notion of an ordinal sum is based on the total order of $[0,1]$ [2]. Together with the fact that the definition of the Archimedean property of t-norms on a bounded poset is not unambiguous (see Section 5), a characterization of continuous (w.r.t. the usual Borel topology) t-norms on $[0,1]^2$ similar to that of continuous t-norms on $[0,1]$ is not to be expected.

Automorphisms of the unit square do not only preserve the direct product structure of t-norms, but clearly also preserve continuity. Related to this, we have formulated the following open problem [4]: *do there exist continuous t-norms on $[0,1]^2$ that are not a direct product of continuous t-norms on $[0,1]$?* To answer this question negatively, one might first try to show that any continuous t-norm on $[0,1]^2$ has $(0,1)$ and $(1,0)$ as idempotent elements.

It is interesting to note that the continuity of a t-norm T on $[0,1]$ is equivalent with stating that

$$(\forall (x,y) \in [0,1]^2)(y \leq x \Rightarrow (\exists z \in [0,1])(T(x,z) = y)),$$

which is nothing else but a reformulation of the intermediate value theorem. This property is also called the *divisibility* of the t-norm T or the *maximal surjectivity* of T [3]. In the case of a finite chain, divisible t-norms are nothing else but the smooth t-norms of Godo and Sierra [7], as shown by Mayor and Torrens [12].

Let us call a t-norm T on $[0,1]^2$ *divisible* if for any x and y in $[0,1]^2$ the following implication holds:

$$y \leq x \Rightarrow (\exists z \in [0,1]^2)(T(x,z) = y).$$

Obviously, the direct product of two strict t-norms on $[0,1]$ is a divisible and cancellative t-norm on $[0,1]^2$. The converse, however, is not clear.

We conclude by formulating the following open problem: *do there exist divisible t-norms on* $[0,1]^2$ *that are not a direct product of continuous (i.e. divisible) t-norms on* $[0,1]$?

Acknowledgement

The first author, Bernard De Baets, is a Post-Doctoral Fellow of the Fund for Scientific Research – Flanders (Belgium). The second author, Radko Mesiar, was partially supported by the grants VEGA 2/6087/99 and GAČR 402/99/0032. The support of COST Action 15 "Many-valued Logics for Computer Science Applications" is also greatly appreciated.

References

[1] G. Birkhoff, *Lattice Theory*, AMS Colloquium Publications, Volume XXV, Providence, Rhode Island, 1967.

[2] A. Clifford, *Naturally totally ordered commutative semigroups*, Amer. J. Math. **76** (1954) 631–646.

[3] B. De Baets, *An order-theoretic approach to solving sup-$\mathcal{T}$ equations*, Fuzzy Set Theory and Advanced Mathematical Applications (D. Ruan, ed.), Kluwer Academic Publishers, 1995, pp. 67–87.

[4] B. De Baets and R. Mesiar, *Triangular norms on product lattices*, Fuzzy Sets and Systems, Special Issue "Triangular norms" (B. De Baets and R. Mesiar, eds.), **104** (1999) 61–75.

[5] G. De Cooman and E. Kerre, *Order norms on bounded partially ordered sets*, J. Fuzzy Math. **2** (1994) 281–310.

[6] C. Drossos and M. Navara, *Generalized t-conorms and closure operators*, Proc. Fourth European Congress on Intelligent Techniques and Soft Computing (Aachen, Germany) (H.-J. Zimmermann, ed.), Vol. 1, ELITE, 1996, pp. 22–26.

[7] L. Godo and C. Sierra, *A new approach to connective generation in the framework of expert systems using fuzzy logic*, Proc. Eighteenth Internat. Symposium on Multiple-Valued Logic (Palma de Mallorca, Spain), IEEE Computer Society Press, 1988, pp. 157–162.

[8] J. Goguen, *L-Fuzzy sets*, J. Math. Anal. Appl. **18** (1967) 145–174.

[9] S. Jenei, *A more efficient method for defining fuzzy connectives*, Fuzzy Sets and Systems **90** (1997) 25–35.

[10] E.-P. Klement, R. Mesiar and E. Pap, *Triangular Norms*, in preparation.

[11] C. Ling, *Representation of associative functions*, Publ. Math. Debrecen **12** (1965) 189–212.

[12] G. Mayor and J. Torrens, *On a class of operators for expert systems*, Internat. J. Intell. Sys. **8** (1993) 771–778.

[13] K. Menger, *Statistical metrics*, Proc. Nat. Acad. Sci. U.S.A. **28** (1942) 535–537.

[14] H. Nguyen and E. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, 1997.

[15] S. Ray, *Representation of a Boolean algebra by its triangular norms*, Mathware & Soft Computing **4** (1997) 63–68.

[16] B. Schweizer and A. Sklar, *Probabilistic Metric Spaces*, Elsevier, 1983.

GENERATED AGGREGATION OPERATORS

Magda Komorníková

STU Bratislava, Radlinského 11, 813 68 Bratislava, Slovakia

and

Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, Pod vodárenskou v í 4, 182 08 Prague, Czech Republic

e-mail magda@vox.svf.stuba.sk

Summary

Aggregation operators generated by means of additive generators are discussed. Depending on the properties of additive generators, some classes of generated aggregation operators are derived. Several examples are included.

Keywords

Additive generator, aggregation operator, quasi-arithmetic mean, triangular norm

1. INTRODUCTION.

Two basic aggregation methods of reals are idempotent ones (min, max, arithmetic mean) and Archimedean ones (sum, bounded sum). Under some additional requirements, these methods can be shown to be the only possible up to isomorphic transformation and composition of operators (e. g. ordinal sum), see for example I-semigroups of Mostert and Shields [15], compare also [9], or quasi-arithmetic means [3]. Because of simplicity, we restrict our considerations to the unit interval inputs and outputs. However, a general real valued scale case is then a matter of rescaling only.

Definition 1. An operator

$$A: \bigcup_{n\in N} [0,1]^n \to [0,1]$$

is called an aggregation operator if it is non-decreasing, it extends the identity on [0, 1], and 0 and 1 are idempotent elements of A, i. e.,

i) $(x_1,\ldots, x_n) \le (y_1, \ldots, y_n) \Rightarrow A(x_1,\ldots, x_n) \le A(y_1, \ldots, y_n)$

ii) $A(x) = x$

and

iii) $A(0, \ldots, 0) = 0$, $A(1, \ldots, 1) = 1$.

Though a general aggregation operator A is, in fact, a collection of n-ary operators, $n \in N$, we will discuss in this contribution binary aggregation operators only (and we will call them also aggregation operators). In the case of associative operators, they fully determin the remaining n-ary operators. In general, the properties of n-ary aggregation operators, $n > 2$, are similar to those for $n = 2$. A large class of aggregation operators can be expressed by means of one-place functions which are called additive generators.

Definition 2. Let f, g: $[0, 1] \to [-\infty, \infty]$ be two continuous non-decreasing functions and let

h: (Ran f + Ran g) $\to$ [0, 1] be another continuous non-decreasing surjective function. Then the binary operator

$$A: [0, 1]^2 \to [0, 1], A(x, y) = h(f(x) + g(y)),$$

is an aggregation operator. The triple (f, g, h) is called a generating triple of A.

Note that A generated by a generating triple (f, g, h) is really a (binary) aggregation operator. More, A is continuous up to the case when the sum $+\infty + (-\infty)$ should be applied (in which case the non-continuity in point (0, 1) or (1, 0) occur). Though several aggregation operators can be described by means of generating triples (weighted quasi-arithmetic means, generated t-norms, t-conorms, uni-norms), the operator min, max, and related operators can never be obtained in such a way [1]. Further, if necessary, the convention fixing the value $+\infty + (-\infty)$ (always either $+\infty$ or $-\infty$) should be given.

2. SYMMETRY.

The symmetry of an operator A generated by a generating triple (f, g, h) is obviously ensured if

$$f(x) + g(y) = f(y) + g(x) \text{ for all } x, y \in [0, 1],$$

i.e., when $g = f + c$, $c = g(0) - f(0)$. However, this condition is not necessary one.

Example 1.

Define

$$f(x) = x,\ g(x) = \begin{cases} \frac{10x^2}{9} & \text{if } x \in [0, 0.9] \\ x & \text{if } x \in]0.9, 1] \end{cases},$$

and $h(x) = \max(0, 10x - 19)$ for $x \in [0, 2]$. Then (f, g, h) is a generating triple generating a symmetric aggregation operator A,

$$A(x, y) = \max(0, 10x + 10y - 19).$$

However, it can be shown that if a generated aggregation operator A is symmetric then it can be generated by some generated triple (f, f, h).

3. NEUTRAL ELEMENT

An aggregation operator A possesses a left unit element $e \in [0, 1]$ if $A(e, x) = x$ for all $x \in [0, 1]$. If A is generated by a generating triple (f, g, h), then $e \in [0, 1]$ is a left unit element of A if and only if the additive generator g is strictly monotone and for all $x \in [f(e) + g(0), f(e) + g(1)]$, $h(x) = g^{-1}(x - f(e))$, i.e.,

$$h(x) = \begin{cases} 0 & \text{if } x \le f(e) + g(0) \\ 1 & \text{if } x \ge f(e) + g(1). \\ g^{-1}(x - f(e)) & \text{otherwise} \end{cases}$$

A similar claim holds in the case of a right unit element. Consequently, we get the following result.

Theorem 1. An element $e \in [0, 1]$ is a unit element of an aggregation operator A generated by a triple (f, g, h) if and only if f is strictly monotone, $g = f + c$ ($c = g(0) - f(0)$), g(e) is finite and

$$h(x) = f^{-1}(\min(f(1), \max(f(0), x - g(e)))),$$

$x \in [f(0) + g(0), f(1) + g(1)]$, i.e., h is the pseudo-inverse of the mapping $f + g(e)$, see [8].

Note that the existence of a neutral element e of a generated aggregation operator A ensures the symmetry of A. More, if $e = 0$ then A is t-conorm (i.e., A is associative), and if $e = 1$, A is a t-norm. Another associative generated aggregation operator A is a uninorm [4, 6] generated by a generating triple (f, f, f^{-1}) where $f: [0, 1] \to [-\infty, \infty]$ is a bijection (hence strictly increasing) and then $e = f^{-1}(0)$.

Example 2 (see also [14]).

Let $f(x) = g(x) = x - e$, $x \in [0, 1]$, $e \in [0, 1]$ and let $h(x) = \min(1, \max(0, x + e))$, $x \in [-2e, 2 - 2e]$. Then the corresponding generated aggregation operator A has neutral element e and is given by

$$A(x, y) = \min(1, \max(0, x + y - e)).$$

Note that A combines the Lukasiewicz t-norm T_L (which corresponds to the case $e = 1$) and t-conorm S_L (corresponding to $e = 0$) in a style described in [6]. However, the resulting operator is associative only if $e = 0$ or $e = 1$.

4. IDEMPOTENCY

The well-known class of idempotent generated aggregation operators is the class of weighted quasi-arithmetic means [3] generated by generating triples (af, bf, h), where f is some strictly monotone additive generator, a, b are positive constants and $h(x) = f^{-1}\left(\frac{x}{a + b}\right)$. Then the corresponding aggregation operator A is given by

$$A(x, y) = f^{-1}\left(\frac{a f(x) + b f(y)}{a + b}\right).$$

However, there are also other idempotent generated aggregation operators.

Theorem 2. An aggregation operator A generated by a generating triple (f, g, h) is idempotent, i.e., $A(x, x) = x$ for all $x \in [0, 1]$ if and only if the sum $f + g$ is strictly monotone and $(f + g)^{-1} = h$.

Example 3.

i) Let $f(x) = x$, $g(x) = x^2$. Then (f, g, h) generates an idempotent aggregation operator A only if

$$h(x) = \frac{1}{2}\left(\sqrt{1 + 4x} - 1\right),\ x \in [0, 2],$$

in which case

$$A(x, y) = \frac{1}{2}\left(\sqrt{1 + 4x + 4y^2} - 1\right);$$

note that A is not symmetric.

ii) Let $f(x) = \max(0, 2x - 1)$, $g(x) = \min(2x, 1)$. Then neither f nor g is strictly monotone. However,

$$f(x) + g(x) = 2x,\ x \in [0, 1],$$

i.e., f + g is strictly monotone.

Consequently, for $h(x)=\frac{x}{2}$, $x \in [0, 2]$, the triple (f, g, h) is a generating triple which generates an idempotent aggregation operator A given by

$$A(x, y) = \max(0, x - \frac{1}{2}) + \min(y, \frac{1}{2}).$$

5. CONCLUSIONS

We have introduced a new class of aggregation operators covering several distinguished classes (e.g., continuous Archimedean t-norms and t-conorms, weighted quasi-arithmetic means, operators discussed in [2, 16], etc.). Several properties of introduced generated aggregation operators were discussed. Further investigations can be done for associative generated aggregation operators (compare approaches given in [2, 4, 6, 13], or for non-continuous generating triples (compare approaches given in [5, 16, 17]). Also limit properties can be of interest (compare [11, 12]) as well as the comparing of generated aggregation operators by means of the corresponding generating triples (compare [7, 10]).

Acknowledgements

The work on this paper was supported by grants GA R 402/99/0032 and VEGA 1/4064/97, 2/6087/99.

References

[1] V.Arnold .(1957) Concerning the representability of functions of two-variables in the form $X[\Phi(x) + \Psi(x)]$. *Uspekhi Mat. Nauk* 12 , pp. 119 – 121

[2] T.Calvo (1998) Comments to [16]. S. Barro et al., eds. *Advances in Fuzzy Logic*, Univ. de Santiago de Compostela, pp. 163 – 177

[3] T.Calvo, G.Mayor and M.Carbonell Extended quasi-linear weighted means. *Int. J. Fuzzy Mathematics,* to appear

[4] T.Calvo., B.De Baets, J.Fodor The functional equations of Frank and Alsina for uninorms and nullnorms. *Fuzzy Sets and Systems*, to appear

[5] J.Fodor, M.Roubens (1994) *Fuzzy Preference Modelling and Multicriteria Decision Support.* Kluwer Academic Publ., Dordrecht

[6] J.Fodor, R.R.Yager, A.Rybalov .(1997) Structure of uninorms. *Internat. J. Uncertain., Fuzziness and Knowledge-Based Systems* 5, pp. 411 – 427

[7] E.P.Klement, R.Mesiar, E.Pap Quasi- and pseudo-inverses of monotone functions, and the construction of t-norms. *Fuzzy Sets and Systems*, in press

[8] E.P.Klement, R.Mesiar, E.Pap (1996) Additive generators of t-norms which are not necessarily continuous. *Proc. EUFIT'96*, Aachen, pp. 70 – 73

[9] E.P.Klement, R.Mesiar, E.Pap (1996) On the relationship of associative compensatory operators to triangular norms and conorms. *Internat. J. Uncertain. Fuzziness and Knowledge-Based Systems* 4, pp. 129 – 144

[10] E.P.Klement, R.Mesiar, E.Pap (1997) A characterization of the ordering of continuous t-norms. *Fuzzy Sets and Systems 86*, pp. 189 – 195

[11] E.P.Klement, R.Mesiar, E.Pap *Triangular norms.* Monograph in preparation

[12] A.Kolesárová (1999) Comparison of quasi-arithmetic means. *Proc. EUROPUSE-SIC'99*, Budapest, pp. 237 - 240

[13] A.Kolesárová Limit theorems for quasi-arithmetic means. *Proc. EUFIT'99*, Aachen, to appear

[14] M.Komorníková, R.Mesiar (1996) Generated aggregation operators for knowledge based systems. *Proc. 3rd CES*, Roma, pp. 519 – 524

[15] M.Mas, G.Mayor, J.Torrens (1999) t-operations. *Internat. J. Uncertain. Fuzziness and Knowledge-Based Systems* 4, pp. 31 – 50

[16] G.Mayor, E. Trillas (1986) On the representation of some aggregation functions. *Proc. ISMVL, IEEE*, pp. 110 – 114

[17] R.Mesiar (1995) Compensatory operators based on triangular norms. *Proc. EUFIT'95*, Aachen, pp. 131 – 135

[18] P.S.Mostert, A.L.Shields (1957) On the structure of semigroups on a compact manifold with boundary. *Ann. of Math.* 65, pp. 117 – 143

[19] P.Viceník (1998) A note on generators of t-norms. *BUSEFAL* 75, pp. 33 – 38

[20] P.Viceník (1998) Additive generators and discontinuity. *BUSEFAL* 76, pp. 25 – 28

SMOOTH ASSOCIATIVE OPERATIONS ON FINITE CHAINS

János Fodor
Department of Biomathematics and Informatics
University of Veterinary Science
István u. 2, H-1078 Budapest, Hungary
E-mail: jfodor@ns.univet.hu

Summary

An intuitive notion of smoothness on finite chains, introduced in [7], is investigated and formulated in a more useful mathematical way. By the help of this equivalent form, we completely characterize the class of smooth associative, increasing binary operations which also satisfy some weak boundary conditions on a chain. Some important subclasses of such operations are also described.

Keywords: ordinal scale, chain, associativity, t-norm, t-conorm, uninorm, nullnorm.

1 INTRODUCTION

One can observe an increasing interest in using linguistic scales (which are totally ordered finite sets or chains in other terminology) in several applications and theoretical problems related to fuzzy sets. Since certain types of associative operations (t-norms, t-conorms, uninorms, nullnorms, etc.) play a key role when the evaluation scale is the closed unit interval, it is important to study associative operations and their relatives (implications for instance) on linguistic scales. Some initial steps have already been made in this direction, see e.g. [7, 9, 3, 10].

The main aim of the present paper is to give the typical (general) form of associative, increasing and smooth binary operations on a finite chain. This study was motivated by a paper of Mayor and Torrens [9]. They characterized all smooth t-norms and t-conorms on finite chains. In some parts of their paper, they used a certain smoothness property borrowed from [7]. This condition is revised in Section 2, where we give a more practical, equivalent form of that smoothness. In Section 3 we determine the typical form of the mentioned operations. Note that such problem was solved by Marichal [8] when the operation was defined on a real interval. We close the paper with short descriptions of some important subclasses.

2 WHAT DOES SMOOTH MEAN ON A FINITE SET?

Assume that $L := \{x_0, x_1, \ldots, x_n, x_{n+1}\}$ is a totally ordered finite set of $n+2$ elements which are indexed increasingly, according to the asymmetric and negatively transitive relation $\prec$: $x_0 \prec x_1 \prec \ldots \prec x_n \prec x_{n+1}$. We use the notation $a := x_0$, $b := x_{n+1}$ in the sequel. For any $x_i, x_j \in L$, $x_i \preceq x_j$ let us define

$$\langle x_i, x_j \rangle := \{x_k \in L \mid x_i \preceq x_k \preceq x_j\},$$

which can be considered as the discrete "closed interval" of points in L between x_i and x_j (note that $x \preceq y$ if and only if either $x \prec y$, or $x = y$).

Mayor and Torrens [9] determined all associative, commutative, increasing binary operations $T : L \times L \to L$ that satisfy $T(b, b) = b$, and for all $x, y \in L$

$$x \preceq y \quad \Longleftrightarrow \quad \exists\, z \in L \text{ such that } x = T(y, z). \quad (1)$$

We might call such a binary operation T a *smooth t-norm on* L, for obvious reasons: it can be seen that T satisfies all the four axioms of t-norms, and condition (1) is just equivalent with continuity of T when it is considered on $[0, 1]$.

In some parts of [9]), condition (1) has been substituted with the following one:

$$T(x_i, x_j) = x_r,\ T(x_{i-1}, x_j) = x_p,\ T(x_i, x_{j-1}) = x_q \text{ imply } r - 1 \leq p, q \leq r. \quad (2)$$

This property was introduced in [7]. It is really a kind of smoothness: if we move only one step (up or down)

from x_i or x_j, while the other argument is unchanged, then the associated function value can move (up or down) at most one step too.

One can easily find increasing operations $M : L \times L \to L$ satisfying associativity, commutativity, $M(b,b) = b$, for which condition (2) holds while (1) does not. Indeed, assume that $\mu \in L$. Let T be a smooth t-norm on $\langle \mu, b \rangle$, and define a binary operation T_μ on L as follows:

$$T_\mu(x,y) := \begin{cases} T(x,y) & \text{if } x,y \in \langle \mu, b \rangle, \\ \mu & \text{otherwise.} \end{cases} \tag{3}$$

Thus defined T_μ is a commutative, associative, increasing operation on L satisfying $T_\mu(b,b) = b$ and condition (2).

Hence, the class of associative, commutative operations M with $M(b,b) = b$ which satisfy condition (2) is wider than the class of smooth t-norms. The main difference is that a smooth t-norm satisfies $T(a,a) = a$ while T_μ does not: $T_\mu(a,a) = \mu$. Hence, for $x, y \in L$, $x \preceq y$ and $x \prec \mu$ there is no $z \in L$ such that $x = T_\mu(y,z)$, since $\mu \preceq T_\mu(y,z)$.

Because of this difference between conditions (1) and (2 for commutative operations, we want to obtain an equivalent, more usable formulation of (2). Since in our approach commutativity of M is not presupposed, we study (2) separately for the two arguments of the operation.

Definition 1. Let M be a binary operation on L. We say that M is *smooth in the first place* if

$$M(x_i,x_j) = x_r,\ M(x_{i-1},x_j) = x_p \quad \text{imply} \quad r-1 \preceq p \preceq r. \tag{4}$$

We call M *smooth in the second place* if

$$M(x_i,x_j) = x_r,\ M(x_i,x_{j-1}) = x_q, \quad \text{imply} \quad r-1 \preceq q \preceq r. \tag{5}$$

□

Now we establish equivalent formulations of (4) and (5) in the following theorem.

Theorem 1. *Let M be a binary operation on L. Condition (4) is equivalent to the following property:*

$$\begin{gathered} M(x_i,x_k) \preceq x_\ell \preceq M(x_j,x_k) \\ \text{if and only if} \\ \exists\, x_m\colon\ x_i \preceq x_m \preceq x_j,\ \text{and } x_\ell = M(x_m,x_k), \end{gathered} \tag{6}$$

and condition (5) has the following equivalent form:

$$\begin{gathered} M(x_k,x_i) \preceq x_\ell \preceq M(x_k,x_j) \\ \text{if and only if} \\ \exists\, x_m\colon\ x_i \preceq x_m \preceq x_j,\ \text{and } x_\ell = M(x_k,x_m) \end{gathered} \tag{7}$$

for all $x_k, x_i, x_j \in L$.

In view of this equivalence, the notion of smoothness of a binary operation M on L means that the so-called *intermediate-value theorem* (see [6], Lemma 1) holds in this finite case too. This makes it possible to borrow easy proof techniques from approaches using topological framework (see [6, 2]).

3 GENERAL FORM OF A SMOOTH ASSOCIATIVE BINARY OPERATION ON L

As it was mentioned, we want to find all binary operations $M : L \times L \to L$ on the finite totally ordered set $L := \{x_0, x_1, \ldots, x_n, x_{n+1}\}$ with ordering $a := x_0 \prec x_1 \prec \ldots \prec x_n \prec x_{n+1} =: b$, which satisfy the following four conditions:

1. M is associative;
2. M is nondecreasing;
3. $M(a,a) = a$, $M(b,b) = b$;
4. M is smooth in the sense of Definition 1.

Denote $\lambda := M(b,a)$, $\mu := M(a,b)$. Since $\lambda, \mu \in L$, there are two cases: either $\lambda \preceq \mu$ or $\mu \preceq \lambda$. We consider only the first case here, the second one can be studied in a similar way. Indeed, if we have an M satisfying the previous conditions, but $\mu \preceq \lambda$, then we can apply the following results for the binary operation $M'(x,y) := M(y,x)$ $(x,y \in L)$.

First of all, some necessary conditions are formulated as a sequence of lemmas. Most of them are almost obviously true. We try to use only those properties of M that we really need in each lemma.

Lemma 1. *Suppose $M : L \times L \to L$ is associative with $M(a,a) = a$, $M(b,b) = b$. Let $\lambda := M(b,a)$, $\mu := M(a,b)$. Then we have*

$$M(\lambda,a) = M(b,\lambda) = \lambda, \tag{8}$$
$$M(a,\mu) = M(\mu,b) = \mu. \tag{9}$$

The next lemma states that associativity and increasingness completely determine values of M, which are constant, in two rectangles around the corners (a,b) and (b,a). In particular, we can see that $M(\lambda,\lambda) = \lambda$ and $M(\mu,\mu) = \mu$.

Lemma 2. *In addition to conditions in Lemma 1, suppose M is increasing and $\lambda \preceq \mu$. Then we have*

$$M(x,y) = \lambda \quad \text{for } x \in \langle \lambda, b \rangle,\ y \in \langle a, \lambda \rangle, \tag{10}$$
$$M(x,y) = \mu \quad \text{for } x \in \langle a, \mu \rangle,\ y \in \langle \mu, b \rangle, \tag{11}$$

By the next statement we can learn in particular that taking the restriction of M to the square $\langle a,\lambda\rangle \times \langle a,\lambda\rangle$, a is the identity of this restriction (like 0 for a t-conorm), while restricting M to $\langle \mu,b\rangle \times \langle \mu,b\rangle$, b is the identity of this second restriction (like 1 for a t-norm).

Lemma 3. *Suppose $M : L \times L \to L$ is associative, increasing and smooth with $M(a,a) = a$, $M(b,b) = b$. Let $\lambda := M(b,a)$, $\mu := M(a,b)$. Then we have*

$$M(a,y) = y \text{ for all } y \in \langle a,\mu\rangle, \quad (12)$$

$$M(b,y) = y \text{ for all } y \in \langle \lambda,b\rangle. \quad (13)$$

We emphasize that symmetry (or commutativity) is not prescribed for M. Nevertheless, on the squares $\langle a,\lambda\rangle \times \langle a,\lambda\rangle$ and $\langle \mu,b\rangle \times \langle \mu,b\rangle$ M turns out to be commutative, as a consequence of the following lemma.

Lemma 4. *Suppose $M : L \times L \to L$ is associative, increasing, smooth, and $M(b,x) = M(x,b) = x$ for all $x \in L$. Then M is commutative on L.*

Thus, applying this lemma to $\langle \mu,b\rangle \times \langle \mu,b\rangle$, and the same for the dual operation on $\langle a,\lambda\rangle \times \langle a,\lambda\rangle$, we can conclude that the previous one is a smooth t-norm, and the second one is a smooth t-conorm. Therefore, the only missing part of $L \times L$ is described now.

Lemma 5. *Suppose $M : L \times L \to L$ is associative, increasing and smooth with $M(a,a) = a$, $M(b,b) = b$. Let $\lambda := M(b,a)$, $\mu := M(a,b)$. Then we have*

$$M(x,y) = y \quad \text{for all } x \in L,\ y \in \langle \lambda,\mu\rangle. \quad (14)$$

We summarize the previous lemmas in the "necessity" part of the following theorem. Before that we introduce S_λ as follows:

$$S_\lambda(x,y) := \begin{cases} S(x,y) & \text{if } x,y \in \langle a,\lambda\rangle, \\ \lambda & \text{otherwise} \end{cases}. \quad (15)$$

Theorem 2. *Consider a function $M : L \times L \to L$ with properties $M(a,a) = a$, $M(b,b) = b$. Let $\lambda := M(b,a)$, $\mu := M(a,b)$, and suppose $\lambda \preceq \mu$. If M is associative, increasing and smooth then M is of the following form $(x,y \in L)$:*

$$M(x,y) = (S_\lambda(x,y) \wedge y) \vee (T_\mu(x,y) \wedge y) \vee (x \wedge y), \quad (16)$$

where functions T_μ, S_λ are defined by (3) and (15), respectively.

The typical form of the operator M is given in Figure 1 when $\lambda \preceq \mu$, and in Figure 2 for the case of $\mu \preceq \lambda$, for sake of simplicity, on the unit square.

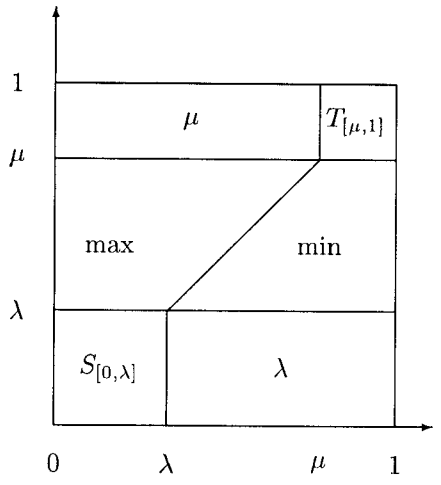


Figure 1: M in case of $\lambda \preceq \mu$

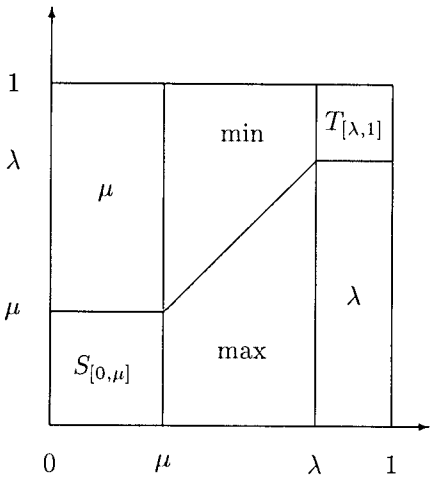


Figure 2: M in case of $\mu \preceq \lambda$

4 PARTICULAR CLASSES

Now we investigate some particular classes of the general case characterized in the previous section. We can recognize subclasses corresponding to well-known families of associative binary operations on $[0,1]$. Denote $\mathcal{C}_{\langle a,b\rangle}$ the class of all associative, increasing, smooth binary operations on $L = \langle a,b\rangle$ (in other words: the class of functions having form (16)).

4.1 COMMUTATIVE OPERATIONS

By Theorem 2, if $M \in \mathcal{C}\langle a,b\rangle$ is commutative then $\lambda = M(b,a) = M(a,b) = \mu$. On the other hand, it is obvious from Eq. (16) that $\lambda = \mu$ is sufficient at the same time to ensure commutativity of $M \in \mathcal{C}\langle a,b\rangle$. Therefore, we obtain that the commutative subclass of

$\mathcal{C}\langle a,b\rangle$ on L is exactly the class of *nullnorms* defined explicitely in [1]. Remark that, in other terminology, this subclass corresponds to associative $T-S$ aggregation functions introduced and investigated in [4]. Note that in [10] so-called *t-operators* have been studied and completely characterized. These operators are exactly the commutative members of $\mathcal{C}\langle a,b\rangle$.

4.2 IDEMPOTENT OPERATIONS

Let $M \in \mathcal{C}\langle a,b\rangle$ be idempotent: $M(x,x) = x$ for all $x \in \langle a,b\rangle$. Then, both $T_{\langle a,\mu\rangle}$ and $S_{\langle\lambda,b\rangle}$ are idempotent, whence $T_{\langle a,\mu\rangle} = \min_{\langle a,\mu\rangle}$ and $S_{\langle\lambda,b\rangle} = \max_{\langle\lambda,b\rangle}$. Consequently, M can be written in the following form:

$$M(x,y) = (\lambda \wedge x) \vee (\mu \wedge y) \vee (x \wedge y) \quad (x,y \in L). \tag{17}$$

This operation was introduced and characterized in [2] as the noncommutative extension of the median (see [6]) in connected order topological spaces. Fortunately, the characterization of (17) is still valid in the present finite ordinal framework.

4.3 OPERATIONS HAVING ABSORBING ELEMENT

Suppose $M \in \mathcal{C}\langle a,b\rangle$ and M has an *absorbing element* $z \in L$: $M(x,z) = M(z,x) = z$ for all $x \in L$. Then, by Lemma 2, we have that $z = \lambda = \mu$, whence M is commutative. Therefore, the subclass $M \in \mathcal{C}\langle a,b\rangle$ with absorbing element is the same as the subclass of commutative operations.

4.4 OPERATIONS HAVING NEUTRAL ELEMENT

Suppose $M \in \mathcal{C}\langle a,b\rangle$ and M has a *neutral element* $e \in L$: $M(x,e) = M(e,x) = x$ for all $x \in L$ (this case corresponds to *uninorms* on $[0,1]$, see [5]). Then, again by Lemma 2, e cannot be strictly between a and b. If $e = a$ then we obtain smooth t-conorms, and if $e = b$ we get smooth t-norms. Therefore, these are the two subclasses of $\mathcal{C}\langle a,b\rangle$ having neutral element. Note that this fact is also proved in [10]. Moreover, similarly to the case of $[0,1]$, pseudo-smooth uninorms on a finite chain are introduced and characterized there.

5 CONCLUSION

In this paper we have studied a class of associative, increasing and smooth binary operations defined on a finite chain. We have shown the general form of such operations, and investigated some particular subclasses too.

Acknowledgements

This work has been supported in part by OTKA T025163.

References

[1] T. Calvo, B. De Baets and J.C. Fodor, The functional equations of Frank and Alsina for uninorms and nullnorms, *Fuzzy Sets and Systems* (to appear).

[2] J.C. Fodor, An extension of Fung–Fu's theorem, *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **4** (1996) 235–243.

[3] J.C. Fodor, Multicriteria decision making without numbers, in: *Proceedings of the International Conference on Methods and Applications of Multicriteria Decision Making* (FUCAM, Mons, 1997), pp. 204–207.

[4] J. Fodor and T. Calvo, Aggregation functions defined by t-norms and t-conorms, in: B. Bouchon-Meunier, Ed., *Aggregation and Fusion of Imperfect Information*, (Physica-Verlag, 1998), pp. 36–48.

[5] J. Fodor, R. Yager and A. Rybalov, Structure of uninorms, *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems* **5** (1997) 411–427.

[6] L.W. Fung and K.S. Fu, An axiomatic approach to rational decision making in a fuzzy environment, in: L.A. Zadeh *et al.*, Eds., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes* (Academic Press, New York, 1975), pp. 227–256.

[7] L. Godó and C. Sierra, A new approach to connective generation in the framework of expert systems using fuzzy logic, in: *Proc. of XVIIIth International Symposium on Multiple-Valued Logic*, Palma de Mallorca, Computer Society Press, Washington D.C. (1988), 157–162.

[8] J.-L. Marichal, On the associativity functional equation, Working Papers G.E.M.M.E. No. 9805, University of Liège, January 1998.

[9] G. Mayor and J. Torrens, On a class of operators for expert systems, *Int. J. of Intelligent Systems* **8** (1993) 771–778.

[10] M. Mas, G. Mayor and J. Torrens, t-operators and uninorms on a finite totally ordered set, *Int. J. of Intelligent Systems* **14** (1999) (to appear).

DISTANCE AGGREGATION OPERATORS

Tomasa Calvo
Department of Mathematics and Computer Science
University of the Balearic Islands
Ctra. de Valldemossa, km 7.5
07071 Palma de Mallorca, Spain.
E-mail: dmitcs0@clust.uib.es

Radko Mesiar
Department of Mathematics
Slovak Technical University
Radlinského 11, 81368 Bratislava, Slovak Republic
Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
P.O. Box 18, 182 08 Praha 8, Czech Republic
E-mail: mesiar@vox.svf.stuba.sk

Summary

Based on the distance of the input set and prescribed point, the class of k–medians of Fung and Fu is recovered. A new class of generalized medians is proposed and its properties are investigated.

Keywords: aggregation operator, distance, median, minimum, maximum, k–median, (a,b)–median.

1 INTRODUCTION

Medians and their generalization play an important role in the aggregation operators theory because of their easy implementation to the ordinal case aggregation. The well–known operators min and max can be understood as medians,

$$\min(x_1,\dots,x_n) = med(x_1,\dots,x_n,\overbrace{0,\dots,0}^{(n-1)-times})$$
$$\max(x_1,\dots,x_n) = med(x_1,\dots,x_n,\overbrace{1,\dots,1}^{(n-1)-times}),$$

where on the right–hand sides new n-1 inputs were added. Several other types of generalized medians can be found in [8, 5].

We will discuss aggregation operators closely related to medians.

Definition 1 Aggregation operator $A : \bigcup_{n\in\mathcal{N}} [0,1]^n \to [0,1]$ is a mapping which is non–decreasing, $A(x) = x$ for each $x \in [0,1]$, and $A(0,\dots,0) = 0$, $A(1,\dots,1) = 1$.

It is easy to see that min is the only aggregation operator A such that for any input we have

$$A(x_1,\dots,x_n) = d(0, A(x_1,\dots,x_n)) \quad (1)$$
$$= d(\{0\},\{x_1,\dots,x_n\}), \quad (2)$$

where d is Euclidean metric on the real line, and for two real subsets $E, F \subseteq \mathbb{R}$,

$$d(E,F) = \inf\{d(x,y) \mid x \in E, y \in F\}.$$

Similarly, max is the only solution of the equation

$$d(1, A(x_1,\dots,x_n)) = d(\{1\},\{x_1,\dots,x_n\}). \quad (3)$$

2 k–MEDIANS

We will generalize (1) and (2) for any given point $k \in [0,1]$. However, requiring (for any input)

$$d(k, A(x_1,\dots,x_n)) = d(\{k\},\{x_1,\dots,x_n\}), \quad (4)$$

for $0 < k < 1$ the resulting operator cannot be an aggregation operator in the sense of Definition 1 (the monotonicity is violated). Therefore we propose a modified version of (3), namely an aggregaion operator A will be called a *k–median* if it fulfills (for any input)

$$d(k, A(x_1,\dots,x_n)) = d(\{k\},[(x_1,\dots x_n)]), \quad (5)$$

where $[(x_1,\dots,x_n)]$ is the smallest interval containing all input values $x_1,\dots,x_n$.

Theorem 1 *For a given $k \in [0,1]$, there is unique k–median V_k and it is an associative idempotent operator introduced by Fung and Fu in 1975 [8].*

Proof. Recall that Fung and Fu [8], see also [6], has introduced the k–medians V_k defined by $V_k = med(x,y,k)$. Due to the associativity of V_k, for any input values $x_1,\dots,x_n$ the resulting aggregation is uniquelly determined and it can be defined by

$$V_k(x_1,\dots,x_n) = med(x_1,\dots,x_n,\overbrace{k,\dots,k}^{(n-1)}), \quad (6)$$

adding new $n-1$ input values k.

It is enough to show that V_k is the unique solution of (4) in the case of two input values x, y. Then:

(i) If $x, y \in [0, k)$, $[(x, y)] = [\min(x, y), \max(x, y)]$ and hence by (4),

$$d(k, A(x, y)) = d(\{k\}, [\min(x, y), \max(x, y)] = k - \max(x, y).$$

Because of $d(k, A(k, k)) = d(\{k\}, [k, k]) = 0$, i.e., $A(k, k) = k$, and the monotonicity of A, it follows $A(x, y) = \max(x, y)$;

(ii) If $x, y \in (k, 1]$, similarly as in i), we can show that $A(x, y) = \min(x, y)$;

(iii) If $x \leq k \leq y$, we have $k \in [(x, y)]$ and hence by (4), $A(x, y) = k$.

Summarizing,

$$A(x, y) = \begin{cases} \max(x, y), & \text{if } x, y < k, \\ \min(x, y), & \text{if } x, y > k, \\ k, & \text{otherwise}, \end{cases}$$

i.e., $A = V_k$, see [7]. Recall that $V_0 = \min$ while $V_1 = \max$. The class $(V_k)_{k \in [0,1]}$ is just the class of all idempotent nullnorms, see [2], or equivalently, the class of all idempotent t-operations, see [10]. ∎

3 (a, b)-MEDIANS

Yager's OWA-operators [11] can be understood as a convex combination of another median-like operators, namely of $W_{n,i} : [0, 1]^n \to [0, 1]$, $n \in \mathcal{N}, i \in \{1, \ldots, n\}$, with

$$W_{n,i}(x_1, \ldots, x_n) = med(x_1, \ldots, x_n, 0, \ldots, 0, 1, \ldots, 1), \tag{7}$$

where $n - i$ new zero inputs and $i - 1$ new one inputs are added. It is easy to see that

$$W_{n,1} = \min \text{ and } W_{n,n} = \max.$$

The next proposed operators generalize both (5) and (6).

Definition 2 For given $a, b \in [0, 1], a \leq b, n \in \mathcal{N}, i \in \{1, \ldots, n\}$, the operator $V_{a,b,n,i} : [0, 1]^n \to [0, 1]$ defined by

$$V_{a,b,n,i}(x_1, \ldots, x_n) = med(x_1, \ldots, x_n, a, \ldots, a, b, \ldots, b), \tag{8}$$

where $n - i$ new a inputs and $i - 1$ new b inputs are added, will be called an (a, b, n, i)-median.

By means of (a, b, n, i)-medians we can define a new class of OWA-like operators, so called MEOWA-operators,

$$MW_{a,b}(x_1, \ldots, x_n) = \sum_{i=1}^{n} w_{n,i} \cdot V_{a,b,n,i}(x_1, \ldots, x_n). \tag{9}$$

Note the $MW_{0,1}$ gives the standard OWA-operators, while $MW_{k,k} = V_k$. More details will be given in a forthcoming paper [4].

Now, we give some properties of (a, b, n, i)-medians which are idempotent, symmetric continuous n-ary aggregation operators:

1. $V_{a,b,n,i} = V_{b,a,n,n-i+1}$,
2. $V_{0,1,n,i} = W_{n,i}$,
3. $V_{k,k,n,i} = V_k / [0, 1]^n$,
4. $V_{a,b,n,1} = V_a / [0, 1]^n$,
5. $V_{a,b,n,n} = V_b / [0, 1]^n$,
6. $d([(a, b)], V_{a,b,n,i}(x_1, \ldots, x_n)) = d([(a, b)], [(x_1, \ldots, x_n)])$.

4 EXAMPLE

Take as an input bag the triple $(0.8, 0.4, 0.6)$. Because of the idempotency of proposed operators, we have always $V_{a,b,3,i} \in [0.4, 0.8]$. The minimal output 0.4 is obtained whenever $a, b \leq 0.4$, or $a \leq 0.4, b \geq 0.4, i = 1$, or $a \geq 0.4, b \leq 0.4, i = 3$. Similarly, the maximal output 0.8 is obtained whenever $a, b \geq 0.8$, or $a \geq 0.8, b \leq 0.8, i = 1$, or $a \leq 0.8, b \geq 0.8, i = 3$. The medain output is obtained whenever $a = b = 0.6$ or $a \leq 0.6, b \geq 0.6, i = 2$. However, any value $c \in [0.4, 0.8]$ can be obtained by means of $V_{a,b,3,i}(0.8, 0.4, 0.6)$, e. g., $V_{c,c,3,i}(0.8, 0.4, 0.6) = V_c(0.8, 0.4, 0.6) = c$, for any $i \in \{1, 2, 3\}$.

5 CONCLUSIONS

We have characterizated by means of distance the class of k-medians. More, inspired by several median-like operators, we have introduced a new class of (a, b, n, i)-medians generalizing both k-medians and basic OWA-operators $W_{n,i}$. Proposed operators can be used for defining a new type of OWA-like operators. More, their median background seems to be promising in the new field of ordinal aggregation operators.

Acknowledgements

The first author has been supported in part by CICYT (TIC96-1393-C06-06), and the second author by VEGA 2/6087/99 and by GACR 402/99/0032.

References

[1] J. Aczél, *On mean values*, Bulletin of the American Math. Society **54** (1948) 392–400.

[2] T. Calvo, B. De Baets and J. Fodor, *The functional equations of Frank and Alsina for uninorms*, J. Uncertain. Fuzziness Knowledge-Based Systems (1999) (To appear).

[3] T. Calvo, B. De Baets and R. Mesiar, *Aggregation operators defined by symmetric k–order additive/maxitive fuzzy measures*, Mathware and Soft Computing (1999) (To appear).

[4] T. Calvo and R. Mesiar, *Generalized medians*, Fuzzy Sets and Systems (In preparation).

[5] J. Fodor, *An extension of Fung-Fu's theorem*, Internat. J. Uncertain. Fuzziness Knowledge-Based Systems **4** (1996) 235–243.

[6] J. Fodor and T. Calvo, *Aggregation functions defined by t-norms and t-conorms*, Aggregation and Fusion of Imperfect Information (B. Bouchon-Meunier, ed.), Physica-Verlag, 1998, pp. 36–48.

[7] J. Fodor, *Aggregation of opinions in a non–homogeneous group*, Proc. IPMU, Paris, 1998, pp. 832–836.

[8] L. Fung and K. Fu, *An axiomatic approach to rational decision-making in a fuzzy environment*, Fuzzy Sets and their Applications to Cognitive and Decision Processes (K. Tanaka, L. Zadeh, K. Fu and M. Shimura, eds.), Academic Press, New York, San Francisco, London, 1975, pp. 227–256.

[9] A. Kolesárová and M. Komorníková, *Triangular norm-based iterative compensatory operators*, Fuzzy Sets and Systems **164** (1999) 109–119.

[10] M. Mas, G. Mayor and J. Torrens, *t–operators*, Int. J. Uncertain Fuzziness Knowledge-Based Systems (1999) (To appear).

[11] R. Yager, *On ordered weighted averaging operators in multicriteria decisionmaking*, IEEE Trans. Systems, Man and Cybernetics **18** (1988) 183–190.

Characterization of commonotone aggregation operators

Pietro Benvenuti, Doretta Vivona, Maria Divari
Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate
Università di Roma "La Sapienza"
Via A. Scarpa, 16, I00161 Roma, Italy
e-mail: benvenut@dmmm.uniroma1.it

Summary

We consider the aggregation operators which are $\oplus$-additive for commonotone functions. The main result is that any operator is expressed by a kind of general fuzzy integral which uses a family of fuzzy measures connected by a $\oplus$-additive Cauchy's equations. In particular, the family of fuzzy measures can be obtained by a $\oplus$-fitting pseudo-multiplication. In this case, the aggregation operator is exactely expressed by a general fuzzy integral [1].

Keywords: Aggregation, Commonotone, Fuzzy integral, Cauchy's equation.

1 INTRODUCTION

Let Ω be an abstract space, $\mathcal{A}$ a σ-algebra of subsets of Ω and $\mathcal{F}$ the family of all $\mathcal{A}$-measurable functions $f : \Omega \to [0, F]$, with $0 < F \leq +\infty$. A chain $\mathcal{M}_f \supset \mathcal{A}$ is associated to any $f \in \mathcal{F}$: it is constituted by the sets

$$C_f(x) =: \{\omega \in \Omega \mid f(\omega) > x\} \qquad x \in [0, F]\,.$$

We suppose that the interval $[0, F]$ has a structure of I-semigroup defined by means of a *pseudo-addition* $\oplus$.

The binary operation $\oplus : [0, F]^2 \to [0, F]$ is called pseudo-addition if it is commutative, associative, monotone non decreasing, continuous, and if 0 is its neutral element.

The structure of the pseudo-additions is known; see [7] or [1] for the general form of the operation $\oplus$. We shall use, as well, a *pseudo-difference* defined for $a < b$ by

$$b \ominus a = \inf\{x \in [0, F] \mid a \oplus x = b\}\,.$$

The family $\mathcal{F}$ of the measurable functions is generated by the set $\mathcal{B}$ of the *basic functions*. For any $a \in [0, F]$ and $A \in \mathcal{A}$, a basic function $b(a, A)$ is defined by:

$$b(a, A)(\omega) = \begin{cases} a & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A. \end{cases} \tag{1}$$

An element s of $\mathcal{F}$ is called *simple function* if its range is finite. The simple functions have several representations by means of basic functions. The well-known classical one is:

$$s = \bigvee_{i=1}^{n} b(a_i, A_i) \tag{2}$$

with $A_i \cap A_j = \phi$ for $i \neq j$ and $0 < a_1 < a_2 < \ldots < a_n$. Moreover, any simple function admits many *$\oplus$-step representations*:

$$s = \bigoplus_{i=1}^{m} b(c_i, C_i) \tag{3}$$

with $C_1 \supseteq C_2 \supseteq \ldots \supseteq C_m$. The standard $\oplus$-step representation is obtained from (2) assuming $m = n$ and:

$$c_1 = a_1\,, \quad c_2 = a_2 \ominus a_1, \ldots, c_n = a_n \ominus a_{n-1}$$
$$C_i = \bigcup_{j=1}^{n} A_j = \{\omega \in \Omega \mid s(\omega) > a_i\}\,.$$

The standard $\oplus$-step representation is minimal for the number of steps and for the heigt of the single steps.

Let $\mathcal{S}$ be the set of all simple functions, we put

$$\mathcal{S}_f =: \{s \in \mathcal{S} \mid s \leq f\}.$$

For every $f \in \mathcal{F}$ it is:

$$f(\omega) = \sup\{s(\omega) \mid s \in \mathcal{S}_f\} \tag{4}$$

because every $f \in \mathcal{F}$ is the limit of a increasing sequence of simple functions ([5] Ch. IV)

It is useful to recall the definition of commonotony for functions [4]. Any function $f : \Omega \to \mathbb{R}$ introduces on Ω a semi-order relation:

$$\omega_1 <_f \omega_2 \iff f(\omega_1) < f(\omega_2)\,.$$

The functions f and g are called *commonotone* ($f \sim g$) if the two corresponding semi-order are non-contradictory, i.e., there exists no pair ω_1, ω_2 in Ω such that $\omega_1 <_f \omega_2$ and $\omega_2 <_g \omega_1$.

We remark that the $\oplus$-step representation of a simple function is built adding commonotone functions. Therefore, the $\oplus$-step representation is called $\oplus$-*commonotone additive representation* too.

2 AGGREGATION OPERATORS

We shall consider the commonotone aggregation operators, i.e., the functionals $L : \mathcal{F} \to [0, F]$ satisfying the following properties:

(L1) $\quad f(\omega) \leq f'(\omega) \Longrightarrow L(f) \leq L(f')$

(monotonicity)

(L2) $\quad f(\omega) = c \quad \forall \omega \in \Omega \Longrightarrow L(f) = c$

(idempotence)

(L3) $\quad f_n \uparrow f \Longrightarrow L(f_n) \uparrow L(f)$

(continuity from below)

(L4) $\quad f_1 \sim f_2 \Longrightarrow L(f_1 \oplus f_2) = L(f_1) \oplus L(f_2)$

(commonotone $\oplus$-additivity).

The properties (L1), (L2), (L3) seem us to be natural for the aggregation operators ([6], [8], [9]) because they are evaluation models or averaging operations. The commonotone $\oplus$-additivity fits the operators to the structure of the semigroup acting on the range of values that shall be aggregated.

For these operators we shall recognize that they are completely determined by their values on the set $\mathcal{B}$, i.e., by the function

$$\ell(a, A) = L[b(a, A)]\,. \tag{5}$$

2.1. LEMMA. *Given a commonotone aggregation operator L, for any $a \in]0, F]$ the map $\mu_a : \mathcal{A} \to [0, F]$ defined by*

$$\mu_a(A) =: \ell(a, A) \tag{6}$$

is a fuzzy measure continuous from below, with $\mu_a(\Omega) = a$.

2.2. LEMMA. For any $\oplus$-step representation (3) of a given simple function s it holds

$$L(s) = \bigoplus_{i=1}^{m} \ell(c_i, C_i)\,. \tag{7}$$

2.3. LEMMA. *For every $f \in \mathcal{F}$ it is:*

$$L(f) = \sup\{L(s) \mid s \in \mathcal{S}_f\}\,. \tag{8}$$

The lemmas above are direct consequences of [(L1)-(L4)]; they show that the family of fuzzy measures $\{\mu_a \mid a \in]0, F]\}$ is associated in biunivocal way to the corresponding operator.

2.4. PROPOSITION. *The family $\{\mu_a\}$, associated to a commonotone aggregation operator, satisfies the following properties:*

(F1) $\quad \mu_a : \mathcal{A} \to [0, F]$ *is a fuzzy measure continuous from below* $\forall\, a \in]0, F]$.

(F2) $\quad \mu_a(\Omega) = a \quad \forall\, a \in]0, F]$

(F3) $\quad a < b \Longrightarrow \mu_a \leq \mu_b$

(F4) $\quad \mu_{a \oplus b} = \mu_a \oplus \mu_b \quad \forall\, a, b \in]0, F]$.

The last property is a link between the measures of every element $A \in \mathcal{A}$. It is expressed by a $\oplus$-additive Cauchy's generalised equation [3].

3 CONSTRUCTION OF COMMONOTONE AGGREGATION OPERATORS

The properties of $\{\mu_a\}$, seen in proposition 2.4, are characteristic of the families of fuzzy measures which correspond to the commonotone $\oplus$-additive operators. This is shown by the following

3.1. THEOREM. *One commonotone aggregation operator enjoing* [(L1)-(L4)] *is associated to every family of fuzzy measures satisfying* [(F1)-(F4)].

The proof of this theorem is laborious but it uses of a quite natural course inspired by the lemmas 2.2 and 2.3. Now we show the line of the demonstration enouncing some lemmas. The proofs of these lemmas are similar to those used in [1] for the construction of the integral with respect to a general fuzzy measure.

Let $\{\mu_a\}$ be a family of fuzzy measures satisfying [(F1)-(F4)].

3.2. LEMMA *For every simple function s the espression*

$$L(s) = \bigoplus_{i=1}^{m} \mu_{c_i}(C_i) \tag{9}$$

has the same value for every its $\oplus$-step representation.

Therefore the expression (9) can be taken as definition of a functional on the set $\mathcal{S}$.

3.3. LEMMA. *The functional* (9) *is non decreasing, continuous from below and $\oplus$-additive.*

3.4. LEMMA. *For every $f \in \mathcal{F}$ the expression* (8) *with $L(s)$ given by* (9) *defines an aggregation operator satisfying the properties* [(L1)-(L4)].

We end the proof of the theorem 3.1 remarking that the construction given by the lemmas 3.2 and 3.4 is the only possible. Every commonotone aggregation operator is a kind of fuzzy general integral built from a suitable family of fuzzy measures.

4 AGGREGATION OPERATORS AS GENERAL INTEGRALS

We can construct a family $\{\mu_a\}$ from a given fuzzy measure m by means of a pseudo-multiplication.

Let $\oplus$ be a given pseudo-addition on $[0, F]$. A binary operation $\odot$: $[0, F] \times [0, M] \to [0, F]$ is called a $\oplus$-fitting *pseudo-multiplication* [1] if the following properties are satisfied:

(M1) $a \le a', \nu \le \nu' \Longrightarrow a \odot \nu \le a' \odot \nu'$

(monotonicity)

(M2) $a \odot 0 = 0 \odot \nu = 0$

(zero element)

(M3) $(\sup a_n) \odot (\sup \nu_m) = \sup(a_n \odot \nu_m)$

(left continuity)

(M4) $(a \oplus b) \odot \nu = (a \odot \nu) \oplus (b \odot \nu)$

(left distributivity).

Let $m : \mathcal{A} \to [0, M]$ $(0 < M \le +\infty)$ be a fuzzy measure continuous from below, we assume:

$$\mu_a(A) = a \odot m(A) \tag{10}$$

4.1. PROPOSITION. *If $m(\Omega)$ is right unitary element for the pseudo-multiplication, the family* (10) *of fuzzy measures satisfies the properties* [(F1)-(F4)].

The corresponding operators is the general integral [2], [1],

$$L(f) = \int^{\oplus} f \odot dm\,. \tag{11}$$

We have the integral form for the aggregation operator if one of the fuzzy measures in the family $\{\mu_a\}$ is discriminant in camparison with the others.

4.2. DEFINITION *The fuzzy measure μ is* discriminant *in the family $\{\mu_a\}$ if*

$$\mu(A) = \mu(A') \Longrightarrow \mu_c(A) = \mu_c(A') \quad \forall c \in [0, F].$$

4.3. PROPOSITION. *If there exists a value $u \in]0, F]$ such that the corresponding fuzzy measure μ_u is discriminant in the family $\{\mu_a\}$, then there exists a $\oplus$-fitting pseudo-multiplication so that*

$$\mu_c(A) = c \odot \mu_u(A)\,. \tag{12}$$

and the expression (11) *holds, with $m = \mu_u$.*

PROOF. From the definition 4.2, $\mu_c(A)$ is a function of c and A, which depends, realy, from c and $\mu_u(A)$:

$$\mu_c(A) = \varphi[c, \mu_u(A)].$$

The operation $\odot$ is defined by putting

$$\begin{cases} 0 \odot m = 0 \\ c \odot m = \varphi[c, m] \quad \forall c \neq 0\,. \end{cases}$$

The pseudo-multiplication satisfies the properties [(M1)-(M4)] and it has u as unit left element, because

$$u \odot \mu_u(A) = \varphi[u, \mu_u(A)] = \mu_u(A)\,.$$

References

[1] P. Benvenuti, R. Mesiar (1999). *Integrals with respect to a general fuzzy measure. Theory and applications.* M. Grabisch, T. Murofushi, M. Sugeno (eds.), Kluwer.

[2] P. Benvenuti, D. Vivona (1996). *General theory of the fuzzy integral.* Mathware and Soft Computing 3 pp 199-209.

[3] P. Benvenuti, D. Vivona, M. Divari. *Cauchy equation with a general pseudo-addition.* Submitted to Aequationes Mathematicae.

[4] D. Dennenberg (1994). *Non-additive Measure and Integral.* Kluwer Academic press.

[5] P. Halmos (1969). *Measure theory.* Van Nostrand Co.

[6] R. Mesiar, M. Komornikova (1997). Proc. Prim. 96, XI Conference on Applied Mathematics - Novi Sad 1997. *Aggregation operators* pp 173-211.

[7] P. S. Mostert, A. L. Shields (1957). *On the structure of semigroups on a compact manifold with boundary.* Ann. of Math. 65 pp 117-143.

[8] R. R. Yager (1994). *Aggregation operators and fuzzy systems modeling.* Fuzzy sets and systems, 67 pp 129-146.

[9] R. R. Yager, A. Rybavov (1996). *Uninorm aggregation operators.* Fuzzy sets and systems, 80 pp 111-120.

(S,U)-integral

Erich Peter Klement
Institute of Mathematics, Johannes Kepler University, 4040 Linz, Austria
e-mail: klement@flll.uni-linz.ac.at

Radko Mesiar
Slovak Technical University, 813 68 Bratislava, Slovakia
and ÚTIA AV ČR, *P.O.Box 18, 182 08 Praha, Czech republic*
e-mail: mesiar@cvt.stuba.sk

Endre Pap
Institute of Mathematics, University of Novi Sad, 21000 Novi Sad, Yugoslavia
e-mail: pap@unsim.ns.ac.yu , pape@eunet.yu

Key words: t-conorm, uninorm, S-measure, (S,U)-integral.

1 Introduction

It is introduced an integral based on t-conorm S, a uninorm U and an S-measure (decomposable measure) with the aim to cover almost all known integrals constructed by pseudo-operations.

A triangular conorm (t-conorm) S is a commutative, associative, monotone binary operation on the unit interval $[0,1]$ with $S(x,0)=x$ for all $x \in [0,1]$. The four basic t-conorms are

$$S_{\mathbf{M}}(x,y)=\max(x,y), \quad S_{\mathbf{P}}(x,y)=x+y-xy,$$

$$S_{\mathbf{L}}(x,y)=\min(x+y,1),$$

$$S_{\mathbf{D}}(x,y)=\begin{cases}\max(x,y) & \text{if } \min(x,y)=0,\\ 1 & \text{otherwise.}\end{cases}$$

2 S-measures

Let X be a fixed non-empty set.

Definition 1 *Let S be a t-conorm and let $\mathcal{A}$ be a σ-algebra of subsets of X. A mapping $m : \mathcal{A} \to [0,1]$ is called an S-measure if it is continuous from below, $m(\emptyset)=0$ and if m is S-decomposable, i.e., for all $A,B \in \mathcal{A}$ with $A \cap B = \emptyset$ we have*

$$m(A \cup B) = S(m(A), m(B)).$$

Remark 2 (i) If S is a left continuous t-conorm, then a set function $m : \mathcal{A} \to [0,1]$ satisfying $m(\emptyset)=0$ is an S-measure if and only if for each sequence $(A_n)_{n\in\mathbf{N}}$ of pairwise disjoint elements of $\mathcal{A}$ we have

$$m\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \mathop{\mathrm{S}}_{n=1}^{\infty} m(A_n).$$

(ii) If the set X is finite or countably infinite and S is a left continuous t-conorm, then each S-measure $m : \mathcal{P}(X) \to [0,1]$ is uniquely determined by the values $m(\{x\})$ with $x \in X$.

(iii) A set function $m : \mathcal{A} \to [0,1]$ is $S_{\mathbf{M}}$-decomposable if and only if for all $A,B \in \mathcal{A}$ we have

$$m(A \cup B) = S_{\mathbf{M}}(m(A), m(B)).$$

Example 3 (i) Obviously, each measure $m : \mathcal{A} \to [0,\infty]$ with $\mathrm{Range}(m) \subseteq [0,1]$ is an $S_{\mathbf{L}}$-measure.

(ii) The function $m : \mathcal{B} \to [0,1]$ defined by $m(A) = \min(\lambda(A),1)$, where $\mathcal{B}$ is the σ-algebra of all Borel subsets of $\mathbb{R}$ and $\lambda : \mathcal{B} \to [0,\infty]$ is the Lebesgue measure, is an $S_{\mathbf{L}}$-measure, but not a measure.

(iii) Fixing a number $K \in \mathbb{N}$ and defining the set function $m : \mathcal{P}(\mathbb{N}) \to [0,1]$ by

$$m(A) = \min\Big(\frac{\mathrm{card}(A)}{K}, 1\Big),$$

where $\mathrm{card}(A)$ denotes the cardinality of the set A, it is easily seen that m is always an $S_{\mathbf{L}}$-measure, and an $S_{\mathbf{D}}$-measure in the case $K=2$, but not a measure.

(iv) The function $m : \mathcal{B} \to [0,1]$ defined by

$$m(A) = 1 - e^{-\lambda(A)}$$

is an $S_{\mathbf{P}}$-measure, for λ as in (ii).

(v) For an arbitrary function $f : X \to [0,1]$, the set function $m_f : \mathcal{P}(X) \to [0,1]$ defined by $m_f(A) = \sup\{f(x) \mid x \in A\}$ is an $S_{\mathbf{M}}$-measure.

(vi) The set function $m : \mathcal{P}([0,1]) \to [0,1]$ defined by

$$m(A) = \begin{cases}\sup A & \text{if } A \text{ is finite or countably}\\ & \text{infinite,}\\ 1 & \text{otherwise,}\end{cases}$$

is an $S_{\mathbf{M}}$-measure.

(vii) Let $P : \mathcal{A} \to [0,1]$ be a probability measure. Then for each $\alpha \in]-1,\infty[\setminus \{0\}$ the set function $m_\alpha : \mathcal{A} \to [0,1]$ defined by

$$m_\alpha(A) = \frac{1}{\alpha}\left((1+\alpha)^{P(A)} - 1\right)$$

is an $S_\alpha^{\mathbf{SW}}$-measure for the Sugeno-Weber t-conorm $S_\alpha^{\mathbf{SW}}(x,y) = \min(x+y+\alpha xy, 1)$. . Observe that we have $m_\lambda = (s_\alpha^{\mathbf{SW}})^{-1} \circ P$, where $s_\alpha^{\mathbf{SW}}$ is the unique additive generator of $S_\alpha^{\mathbf{SW}}$ which satisfies $s_\alpha^{\mathbf{SW}}(1) = 1$ (compare [16]). We remark that m_λ is subadditive if and only if $\lambda \in]-1,0[$, and superadditive if and only if $\lambda \in]0,\infty[$.

(viii) If $m : \mathcal{A} \to [0,1]$ is an S-measure and $\psi : [0,1] \to [0,1]$ an increasing bijection, then $\psi^{-1} \circ m : \mathcal{A} \to [0,1]$ is an S_ψ-measure, where S_ψ is defined by

$$S_\psi(x,y) = \psi^{-1}\left(S(\psi(x),\psi(y))\right).$$

The S-measures presented in Example 3(i)–(iv),(vii) are special cases of *(σ-)S-decomposable measures* [1, 18], some of which can be constructed as follows:

Proposition 4 *Let S be a continuous Archimedean t-conorm, $s : [0,1] \to [0,\infty]$ an additive generator of S, $\mathcal{A}$ a σ-algebra of subsets of X and $m : \mathcal{A} \to [0,\infty]$ a measure. Then the function $s^{(-1)} \circ m : \mathcal{A} \to [0,1]$ is an S-measure (here $s^{(-1)}$ is the pseudo-inverse of s).*

Definition 5 *Let S be a t-conorm, $\mathcal{A}$ a σ-algebra of subsets of X and $m : \mathcal{A} \to [0,1]$ an S-measure. A set $A \in \mathcal{A}$ is called S-m-faithful if for each $B \in \mathcal{A}$ with $B \subseteq A$ such that, whenever $(u,v) \in [0, m(B)[\times[0, m(A \setminus B)[$, we have $S(u,v) < 1$. A partition $\mathcal{C} = \{C_k \mid C_k \in \mathcal{A}, k \in K\}$ of X, where K is finite or countably infinite , is called an S-m-partition if, for each $k \in K$, the set C_k is S-m-faithful. The S-measure is said to be S-faithful if there exists an S-m-faithful partition $\mathcal{C}$ of X.*

Remark 6 Let S be a t-conorm, $\mathcal{A}$ a σ-algebra of subsets of X and $m : \mathcal{A} \to [0,1]$ an S-measure.
(i) If $A \in \mathcal{A}$ is S-m-faithful then each $B \in \mathcal{A}$ with $B \subseteq A$ is S-m-faithful.
(ii) Each $A \in \mathcal{A}$ with $m(A) < 1$ is S-m-faithful.
(iii) If we have $S(x,y) = 1$ if and only if $\max(x,y) = 1$ then X is S-m-faithful. Consequently, each $S_{\mathbf{M}}$-measure is $S_{\mathbf{M}}$-faithful.
(iv) If S is a nilpotent t-conorm, then the S-faithfulness of an S-measure is a weaker property than the m-achievability of X, see [18].

An interesting relationship between (σ-additive) measures and S-measures with respect to some continuous Archimedean t-conorm S is given as follows:

Proposition 7 *Let S be a continuous Archimedean t-conorm, $s : [0,1] \to [0,\infty]$ an additive generator of S, $\mathcal{A}$ a σ-algebra of subsets of X and $m : \mathcal{A} \to [0,1]$ an S-measure. Then $s \circ m : \mathcal{A} \to [0,\infty]$ is a measure if and only if X is S-m-faithful.*

The assumption that X is S-m-faithful cannot be dropped because of Example 3 (ii),(iii). Observe that, in (ii), a set $A \in \mathcal{B}$ is $S_{\mathbf{L}}$-m-faithful if and only if $\lambda(A) \leq 1$, and in (iii), if and only if $\mathrm{card}(A) \leq K$.

3 (S,U)-integral

A uninorm U (respectively t-norm T) is a commutative, associative, monotone binary operation on the unit interval $[0,1]$ and for some $0 < e < 1$ we have $U(x,e) = x$ (respectively $T(x,1) = x$) for all $x \in [0,1]$. Throughout this section, let U be a left-continuous uninorm or t-norm and S a continuous t-conorm such that U is conditionally distributive over S, i.e., they satisfy the property (CD):

$$U(x,S(y,z)) = S(U(x,y),U(x,z))$$

for all $x,y,z \in [0,1]$ such that $S(y,z) < 1$.

Denote by $\mathcal{M}$ the set of all $\mathcal{A}$-measurable functions from X to $[0,1]$. As usual, a measurable step function $\varphi : X \to [0,1]$ is a measurable function which assumes only finitely many values, and the set of all step functions will be denoted $\mathcal{S}$. If $\mathrm{Range}(\varphi) = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ with $\alpha_i \neq \alpha_j$ whenever $i \neq j$, and if $A_i = \varphi^{-1}(\{\alpha_i\})$, then there is a canonical representation of φ given by

$$\varphi = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}. \qquad (1)$$

Observe that representation (1) is independent of the t-conorm S under consideration. Let $m : \mathcal{A} \to [0,1]$ be an S-faithfull S-measure in whole section.

Definition 8 *(i) The (S,U)-integral of a measurable step function $\varphi : X \to [0,1]$ (which is represented as in (1)) is defined by*

$$\int_X^{(S,U)} \varphi \, dm = \mathop{\mathrm{S}}_{k\in K} \Big(\mathop{\mathrm{S}}_{i=1}^{n} U(\alpha_i, m(A_i \cap C_k)) \Big),$$

where $\mathcal{C}$ is an S-m-partition.

(ii) The (S,U)-integral of a measurable function $f : X \to [0,1]$ is defined by

$$\int_X^{(S,U)} f \, dm = \sup \Big\{ \int_X^{(S,U)} \varphi \, dm \mid \varphi \in \mathcal{S}, \varphi \le f \Big\}.$$

(iii) The (S,U)-integral of a measurable function $f : X \to [0,1]$ on a set $A \in \mathcal{A}$ is defined by

$$\int_A^{(S,U)} f \, dm = \int_X^{(S,U)} f_A \, dm,$$

where the function f_A is equal f on the set A and otherwise it is zero.

Theorem 9 *For all $f, h \in \mathcal{M}$ we have*

$$\int_X^{(S,U)} f \, dm \le \int_X^{(S,U)} h \, dm$$

whenever $f \le h$,

$$\int_X^{(S,U)} S(f,h) \, dm = S\Big(\int_X^{(S,U)} f \, dm, \int_X^{(S,U)} h \, dm \Big),$$

whenever for all $a \in [0,1]$ we have $U(S(f,h),a) = S(U(f,a),U(h,a))$.

Theorem 10 *For each $f \in \mathcal{M}$, the function $m_f : \mathcal{A} \to [0,1]$ defined by*

$$m_f(A) = \int_A^{(S,U)} f \, dm,$$

is an S-faithful S-measure.

On the family

$$\{ f \in \mathcal{M} \mid X \text{ is } S - m_f \text{ -faithful} \}$$

the (S,U)-integral is U-homogeneous.

For the (S,U)-integral some classical convergence theorems such as the monotone convergence theorem and Fatou's lemma, can be proven, as well as a Radon-Nikodým theorem for the (S,U)-integral [4] (compare also [5, 13, 17]).

For special uninorms or t-norms U and t-conorms S, we obtain some integrals which are well-known in the literature.

Example 11 (i) Let $m : \mathcal{A} \to [0,1]$ be an $S_{\mathbf{L}}$-measure which is also a (σ-additive) measure. Then the $(S_{\mathbf{L}}, T_{\mathbf{P}})$-integral of a function $f \in \mathcal{M}$ coincides with the classical Lebesgue integral of f.

(ii) For each left-continuous U, for each $S_{\mathbf{M}}$-measure $m : \mathcal{A} \to [0,1]$ and for each function $f \in \mathcal{M}$ we have that $(S_{\mathbf{M}}, U)$-integral is given by

$$\sup \big\{ U(a, m(\{x \in X \mid f(x) \ge a\})) \;\big|\; a \in [0,1] \big\}.$$

If m is a completely maxitive measure then we have

$$\int_X^{(S_{\mathbf{M}},U)} f \, dm = \sup\{ U(f(x), m(\{x\})) \mid x \in X \}.$$

In particular, the $(S_{\mathbf{M}}, T_{\mathbf{M}})$-integral coincides with the integral introduced in [16], and the $(S_{\mathbf{M}}, T_{\mathbf{P}})$-integral coincides with the integral studied in [15].

(iii) The restriction to S-measures and measurable functions whose ranges are subsets of $[0,1]$ causes no loss of generality. Let $([a,b], \oplus, \odot)$, for $[a,b] \subseteq [-\infty, \infty]$, be a continuous semiring, see [2], with zero element a (see [3]). Consequently, by [3], there is some strictly increasing bijection $\psi : [a,b] \to [0,1]$ such that the continuous t-conorm S defined by

$$S(x,y) = \psi(\psi^{-1}(x) \oplus \psi^{-1}(y))$$

is isomorphic to $\oplus$. Define the operation (uninorm or t-norm) U on $[0,1]$ by

$$U(x,y) = \psi(\psi^{-1}(x) \odot \psi^{-1}(y))$$

and observe that U is conditionally distributive over S. Then for an $\mathcal{A}$-measurable function $f : X \to [a,b]$ and an $\oplus$-faithful $\oplus$-measure $m : \mathcal{A} \to [a,b]$ one can introduce $(\oplus, \odot)$-integral in complete analogy to our construction. Such integral was introduced in [8, 11] and can be represented by the (S,U)-integral via

$$\int_X^{(\oplus,\odot)} f \odot dm = \psi^{-1}\Big(\int_X^{(S,U)} (\psi \circ f) \, d(\psi \circ m) \Big).$$

(iv) For $[a,b] = [0,\infty]$ we obtain an integral introduced in [17], and for $([a,b], \oplus, \odot) = ([0,\infty], +, \cdot)$ we again come back to the classical Lebesgue integral.

(v) If the operations $\oplus$ and $\odot$ in the semiring $([a,b], \oplus, \odot)$ are generated by some strictly

increasing bijection $g : [a, b] \to [0, \infty]$ via

$$\begin{aligned} x \oplus y &= g^{-1}(g(x) + g(y)), \\ x \odot y &= g^{-1}(g(x)g(y)), \end{aligned}$$

the integral introduced in [11] (called g-integral in [7, 12]) has the special form

$$\int_X^{(\oplus,\odot)} f \odot dm = g^{-1}\Big(\int_X (g \circ f)\, d(g \circ m)\Big).$$

Under some additional conditions, also a g-derivative can be defined [12] which turns out to be useful when solving some differential equations (see [8, 13]).

(vi) If S is a continuous Archimedean t-conorm with an additive generator $s : [0, 1] \to [0, \infty]$ and if we replace U by the binary operation $*$ given by

$$a * b = s^{-1}(a \cdot s(b))$$

we can repeat the whole construction of Definition 8 for the $(S, *)$-integral, and we obtain the integral introduced in [18].

Further generalizations and modifications of the (S, U)-integral can be found in [5, 6, 8, 9, 10, 13, 14].

Acknowledgement

The third author would like to thank the Department of Mathematics of the Johannes Kepler University, Linz, Austria, for the financial support of their visits. The second author was supported by grants VEGA 2/6087/99, GACR 402/99/032 and the first author with the second author were supported also by Action Austria-Slovakia 26s34.

References

[1] D. Dubois, H. Prade, A class of fuzzy measures based on triangular norms, Internat. J. Gen. System 8 (1982), 43–61.

[2] J.S. Golan, *The theory of semirings with applications in mathematics and theoretical computer science*, Longman, 1992.

[3] E.P. Klement, R. Mesiar, E. Pap, *Triangular Norms* (to appear).

[4] E.P. Klement, R. Mesiar, E. Pap, (S, U)-integral based on S-measure (preprint).

[5] E.P. Klement, S. Weber, Generalized measures, Fuzzy Sets and Systems 40 (1991), 375–394.

[6] A. Kolesárová, A note to the $\oplus$-measure based integrals, Tatra Mt. Math. Publ. 3 (1993), 173-182.

[7] A. Marková, A note on g-derivative and g-integral, Tatra Mt. Math. Publ. 8 (1996), 71-76.

[8] V.P. Maslov, S.N. Samborskij (eds.), *Idempotent Analysis*, Advances in Soviet Mathematics 13, Amer. Math. Soc., Providence, Rhode Island, 1992.

[9] R. Mesiar, Choquet–like integrals, J. Math. Anal. Appl, 194 (1995), 477–488.

[10] T. Murofushi, M. Sugeno, Fuzzy t–conorm integrals with respect to fuzzy measures: generalization of Sugeno integral and Choquet integral, Fuzzy Sets and Systems 42 (1991), 57–71.

[11] E. Pap, Integral generated by decomposable measure, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 20,1 (1990), 135-144.

[12] E. Pap, $g-$calculus, Univ. u Novom Sadu Zb. Rad. Prirod.-Mat. Fak. Ser. Mat. 23,1 (1993), 145-156 .

[13] E.Pap, *Null-Additive Set Functions*, Kluwer Academic Publishers, Dordrecht-Boston-London, 1995.

[14] M. Sion, *A Theory of Semigroup Valued Measures*, Springer-Verlag, 1973.

[15] N. Shilkret, Maxitive measure and integration, Indag. Math. 3 (1971) (1971), 109-116.

[16] M. Sugeno, Theory of fuzzy integrals and its applications, Ph. D. Thesis, Institute of Technology, Tokyo, 1974.

[17] M. Sugeno, T. Murofushi, Pseudo-additive measures and integrals, J. Math. Anal. Appl. 122 (1987), 197-222.

[18] S. Weber, $\perp$ $-$decomposable measure and integrals for Archimedean $t-$conorms, J. Math. Anal. Appl. 101 (1984), 114-138.

CONVEXITY OF FUZZY COALITION GAMES*

Milan Mareš

ÚTIA AV ČR
P. O. Box 18, 182 08 Praha 8, Czech Republic
mares@utia.cas.cz

Summary

The deterministic coalition game theory is based on the presumption that all expected incomes of coalitions and individual players are exactly known even befor the bargaining process. If we leave this presumption then we extend the original game model into the fuzzy coalition game with vague, i.e., fuzzy, pay-offs. The vagueness of the pay-offs leads to vagueness of some other concepts and properties like the superadditivity and convexity of the game or its core. In this note we formulate the concept of the convexity of fuzzy coalition games and discuss its properties including the convexity of the game and the non-emptiness of its core, known from the deterministic coalition games theory. The type of coalition games dealt in this contribution are the games with side-payments.

Keywords: Game, Coalition game, Core of coalition game, Fuzzy quantity, Fuzzy coalition game, Convex game

1 COALITION GAME WITH SIDE–PAYMENTS

The model of the coalition game with side-payments whose fuzzified modification is investigated here is dealt in many works. Let us remember, as a representation of them, e. g., [8] and [9]. Due to this model, every *coalition game with side-payments* is defined as a pair (I, v), where I is the non-empty and (usually) finite *set of players* and $v : 2^I \to R$, a mapping connecting each set $K \subset I$ with a real number $v(K)$, is called *characteristic function* of the game. Subsets K of I are called *coalitions*. It is natural to put $v(\emptyset) = 0$. Value $v(K)$ represents the expected income of the coalition K.

We say that the game (I, v) is *superadditive* iff for every pair of disjoint coalitions K, $L \subset I$

$$v(K \cup L) \geq v(K) + v(L).$$

Any real-valued vector $\boldsymbol{x} = (x_i)_{i \in I} \in R^I$ is called an *imputation*. The set of imputations C such that

$$C = \Big\{ \boldsymbol{x} \in R^I : \textstyle\sum_I x_i \leq v(I) \\ \text{and } \textstyle\sum_K x_i \geq v(K) \text{ for all } K \subset I \Big\} \tag{1}$$

is called the *core of the game* (I, v).

The core, generally, need not be non-empty. A well known sufficient condition for its non-emptiness is the convexity. We say, that (I, v) is *convex* iff for every pair of coalitions K, $L \subset I$

$$v(K \cup L) + v(K \cap L) \geq v(K) + v(L). \tag{2}$$

It is easy to see that the convexity implies the superadditivity of the game.

Due to a well known result, remembered, e. g., in [8, 9], every convex coalition game with side-payments has non-empty core (1).

2 FUZZY QUANTITY

The main fuzzy set theoretical tool used in the following sections for the fuzzification of the coalition game model is the concept of fuzzy quantity. Let us remember, at least briefly, those of its properties used in this paper.

Every fuzzy subset a of the real line R with membership function $\mu_a : R \to [0, 1]$ such that

$$\exists\, x_a \in R : \mu_a(x_a) = 1, \tag{3}$$

$$\exists\, x_1, x_2 \in R : x_1 < x_2,\ \mu_a(x) = 0 \quad \text{for all } x \notin [x_1, x_2]. \tag{4}$$

* Particular research summarized in this paper was partly supported by the Key-Project of the Academy of Sciences of the Czech Republic No. K 1075601, by the grant of the Grant Agency of the Academy of Sciences No. A 1075905, grant of the Grant Agency of the Czech Republic No. 402/99/0032, and by the Project of the Ministry of Education, Youth and Sports of the Czech Republic project No. VS 96063.

is called *fuzzy quantity* and the number x_a in (3) is its *modal value*. The fuzzy quantities can be processed by algebraical methods summarized, e. g., in [2] and [4]. In this paper we need to remember the operation of summation and the ordering relation.

Let a, b be fuzzy quantities with membership functions μ_a, μ_b, respectively. Then the sum $a \oplus b$ is also a fuzzy quantity and its membership function $\mu_{a\oplus b}$ is defined by

$$\mu_{a\oplus b}(x) = \sup_{y\in R} \left(\min(\mu_a(y),\ \mu_b(x-y))\right), \quad x \in R. \tag{5}$$

There exist many approaches to the ordering relation over fuzzy quantities which can be found in the accessible literature (e. g., in [3] and [2, 4]). Here, we use the one which is based on the presumption that relation between vague (fuzzy) objects is necessarily also vague being valid with some possibility. If a, b are fuzzy quantities then the possibility that they fulfil weak ordering relation $a \succsim b$ is denoted by $\nu_{\succsim}(a, b) \in [0, 1]$ and defined by

$$\nu_{\succsim}(a, b) = \sup_{\substack{x,y\in R \\ x\geq y}} \left(\min(\mu_a(x),\ \mu_b(y))\right). \tag{6}$$

It is evident that $\nu_{\succsim} = 1$ if $x_a \geq x_b$ and that for any real number $r \in R$ the relations $r \succsim a$ and $a \succsim r$ are valid with possibilities

$\nu_{\succsim}(r, a) = \sup(\mu_a(x) : x \in R,\ r \geq x)$,

$\nu_{\succsim}(a, r) = \sup(\mu_a(x) : x \in R,\ x \geq r)$.

3 FUZZY EXTENSION OF COALITION GAME

Let (I, v) be a coalition game and let us suppose that the values $v(K)$, $K \subset I$, are not exactly determined in the time of bargaining and coalition forming. Then it is much more realistic to substitute deterministic values $v(K)$ by some fuzzy quantities reflecting the vagueness of the expectations of the results of cooperation. For every $K \subset I$ we define fuzzy quantity $w(K)$ with membership function $\mu_K : R \to [0, 1]$, where for every $K \subset I$, $K \neq \emptyset$,

$$\mu_k(v(K)) = 1, \tag{7}$$

i. e., $v(K)$ is the modal value of $w(K)$,

$$\mu(K) \text{ is increasing for } x < v(K) \text{ and decreasing for } x > v(K), \tag{8}$$

and for empty coalition $K = \emptyset$

$$\begin{aligned} \mu_\emptyset(x) &= 1 \quad \text{if} \quad x = v(K), \\ &= 0 \qquad x \neq v(K). \end{aligned} \tag{9}$$

The pair (I, w) is called *fuzzy extension* of the game (I, v) or *fuzzy coalition game*.

Some properties of (I, w) are investigated in [5, 6, 7]. As the expected values of incomes are vague, some other concepts derived from them have to be vague, as well. It regards also the core and the convexity concept.

The core of (I, w) is defined as fuzzy subset C_F of R^I with membership function $\gamma_C : R^I \to [0, 1]$, such that for $\boldsymbol{x} = (x_i)_{i\in I} \in R^I$

$$\gamma_C(\boldsymbol{x}) = \min_{K\subset I} \left(\nu_{\succsim}(w(I), \textstyle\sum_I x_i),\ \nu_{\succsim}(\sum_K x_i,\ w(K)) \right). \tag{10}$$

The possibility that there exists a non-empty core C_F in the fuzzy game (I, w) will be denoted by

$$\gamma_C(I, w) = \sup\left(\gamma_C(\boldsymbol{x}) : \boldsymbol{x} \in R^I\right). \tag{11}$$

4 FUZZY SUPERADDITIVITY AND CONVEXITY

The important properties of superadditivity and convexity can be extended to fuzzy coalition games. We say that (I, w) is *fuzzy superadditive* if for any pair of disjoint coalitions K, $L \subset I$

$$w(K \cup L) \succsim w(K) \oplus w(L). \tag{12}$$

Relation (12) for a given pair K, L is valid with possibility

$$\overline{\nu}(K, L) = \nu_{\succsim}(w(K \cup L),\ w(K) \oplus w(L)) \tag{13}$$

due to (6) and, consequently, the possibility that (I, w) is fuzzy superadditive is

$$\sigma(I, w) = \min\left(\overline{\nu}(K, L) : K, L \subset I,\ K \cap L = \emptyset\right). \tag{14}$$

Analogously, we say that (I, w) is *fuzzy convex* iff for any pair of coalitions K, $L \subset I$

$$w(K \cup L) \oplus w(K \cap L) \succsim w(K) \oplus w(L) \tag{15}$$

which relation is valid with possibility

$$\nu^*(K, L) = \nu_{\succsim}(w(K \cup L) \oplus w(K \cap L),\ w(K) \oplus w(L)). \tag{16}$$

due to (6), and, consequently, the possibility that (I, w) is fuzzy convex is

$$\delta(I, w) = \min\left(\nu^*(K, L) : K, L \subset I\right). \tag{17}$$

Proposition 1. It is evident that for disjoint K, $L \subset I$ the equality $\overline{\nu}(K, L) = \nu^*(K, L)$ holds and, consequently, $\delta(I, w) \leq \sigma(I, w)$.

Proof. The statement follows from definitions, immediately. □

Proposition 2. If (I, v) is convex then $\gamma_C(I, w) = 1$.

Proof. The convexity of (I, v) means that there exists its non-empty core C and this, due to (6), (10), (11) implies $\gamma_C(I, w) = 1$. □

The previous result can be extended on more general cases.

Proposition 3. If (I, w) is fuzzy extension of (I, v) with $\mu_K(x)$ strictly increasing for $x < v(K)$ and strictly decreasing for $x > v(K)$, for any $K \subset I$, and x such that $\mu_K(x) \neq 0$, then

$$\gamma_C(I, w) \geq \delta(I, w).$$

Proof. Let (I, w) be fuzzy convex with possibility $\delta(I, w)$. Then there exist real values $v^*(K)$ for each $K \subset I$, such that $\mu_K(v^*(K)) \in [0, 1]$,

$$\min(v^*(K) : K \subset I) = \delta(I, w)$$

and the deterministic game (I, v^*) is convex. It means that (I, v^*) has a non-empty core C^*. As the fuzzy coalition game (I, w) earns the values of the deterministic game (I, v^*) with the possibility $\delta(I, w)$ then the set C^* is a subset of C_F at least with the same possibility and, consequently, $\gamma_C(I, w)$ is at least equal to $\delta(I, w)$. □

5 AN EXAMPLE

To illustrate the above concepts and propositions we can consider a coalition game with side-payments (I, v) with 3 players such that $I = \{1, 2, 3\}$, and for all $i, j = 1, 2, 3$, $i \neq j$, $v(\{i\}) = 1/5$, $v(\{i, j\}) = 2$, $v(I) = 2$. The game (I, v) is neither superadditive (e. g. $v(\{1\}) + v(\{2, 3\}) > v(I)$), nor convex. Its deterministic core $C = \emptyset$. Let us consider fuzzy extension (I, w) of (I, v) such that for $i, j \in I$, $i \neq j$, $x \in R$,

$$\begin{aligned} \mu_{\{i\}}(x) &= x + 4/5 \quad \text{for} \quad x \in [-4/5, 1/5], \\ &= 6/5 - x \quad \text{for} \quad x \in [1/5, 6/5], \\ &= 0 \qquad \text{else}, \end{aligned}$$

$$\begin{aligned} \mu_{\{i,j\}}(x) = \mu_I(x) &= x - 1 \quad \text{for} \quad x \in [1, 2], \\ &= 3 - x \quad \text{for} \quad x \in [2, 3], \\ &= 0 \qquad \text{else}. \end{aligned}$$

Then it is easy to verify that for $i, j, k \in I$, $i \neq j \neq k \neq i$

$$\underset{\lesssim}{\nu}(w(\{i, j\}), w(\{i\}) \oplus w(\{j\})) = 1$$

$$\underset{\lesssim}{\nu}(w(I), w(\{i, j\}) \oplus w(\{k\})) = 14/15$$

and, consequently, (I, w) is superadditive with possibility

$$\sigma(I, w) = 14/15.$$

analogously, for $i, j, k \in I$, $i \neq j \neq k \neq i$,

$$\underset{\lesssim}{\nu}(w(I) \oplus w(\{j\}), w(\{i, j\}) \oplus w(\{j, k\})) = 11/20$$

where, evidently, the membership functions of $w(I) \oplus w(\{j\})$ and $w(\{i, j\}) \oplus w(\{j, k\})$ are $\mu_{I,j}$, $\mu_{(i,j)(j,k)}$, respectively, where

$$\begin{aligned} \mu_{I,j}(x) &= (5x - 1)/10 \quad \text{for} \quad x \in [1/5, 11/5], \\ &= (21 - 5x)/10 \quad \text{for} \quad x \in [11/5, 21/5], \\ &= 0 \qquad \text{else}, \end{aligned}$$

$$\begin{aligned} \mu_{(i,j)(j,k)}(x) &= (x - 2)/2 \quad \text{for} \quad x \in [2, 4], \\ &= (6 - x)/2 \quad \text{for} \quad x \in [4, 6], \\ &= 0 \qquad \text{else}. \end{aligned}$$

It means that (I, w) is fuzzy convex with possibility

$$\delta(I, w) = 11/20.$$

Finally, the fuzzy core C_F of (I, w) is non-empty with possibility

$$\gamma(I, w) = 3/5.$$

This value is achieved for the imputation $\boldsymbol{x} = (4/5, 4/5, 4/5)$ which can be realized by the coalition I with the possibility $\mu_I(12/5) = 3/5$ as

$$\underset{\lesssim}{\nu}(w(I), 12/5) = \mu_I(12/5),$$

and for any other $\boldsymbol{y} \in R^3$ either $\mu_I(\boldsymbol{y}) < 3/5$ if $\boldsymbol{y} > \boldsymbol{x}$. or for some 2-players coalition $\{i, j\}$

$$\begin{aligned} &\underset{\lesssim}{\nu}(y_i + y_i, w(\{i, j\})) \\ < \ &\underset{\lesssim}{\nu}(3/5 + 3/5, w(\{i, j\})) = \mu_{\{i,j\}}(8/5) = 3/5. \end{aligned}$$

Note, that

$$\gamma(I, w) \geq \delta(I, w).$$

References

[1] D. Butnariu, E. P. Klement (1993). *Triangular Norm–Based Measures and Games With Fuzzy Coalitions*. Kluwer, Dordrecht.

[2] D. Dubois, H. Prade (1998). Fuzzy numbers: An overview. In: *Analysis of Fuzzy Information* (J.C Bezdek - Ed.). CRC–Press, Boca Raton 1988, Vol. 2, pp 3–39.

[3] E. E. Kerre, X. Wang (1999). Reasonable properties for the ordering of fuzzy quantities. Part I, Part II. *Fuzzy Sets and Systems*. To appear.

[4] M. Mareš (1994). *Computation Over Fuzzy Quantities*. CRC–Press, Boca Raton.

[5] M. Mareš (1999). Superadditivity in fuzzy extensions of coalition games. *Tatra Mountains Math. Publ.* To appear.

[6] M. Mareš (1999). Fuzzy coalition structures. *Fuzzy Sets and Systems*. To appear.

[7] M. Mareš (1995). Coalition forming motivated by vague profits. In: *Proceedings of MME'95*, Ostrava, TU of Ostrava 1995, pp 114–199.

[8] J. Rosenmüller (1972). *Kooperative Spiele und Märkte*. Springer Verlag, Heidelberg.

[9] J. Rosenmüller (1982). *The Theory of Games and Markets*. North Holland, Amsterdam.

BALANCING PROPERTY OF FUZZY MEASURES

T. Calvo[1], J. Martín[2], G. Mayor[3] and J. Torrens[4]

Department of Mathematics and Computer Science
University of the Balearic Islands
Ctra. de Valldemossa, km 7.5
07071 *Palma de Mallorca, Spain.*

[1]dmitcs0@clust.uib.es [2]javier.martin@uib.es [3]dmigmf0@clust.uib.es [4]dmijts0@clust.uib.es

Summary

In this work, we first introduce a balancing property on a fuzzy measure. After that, we give the conditions of when an additive fuzzy measure satisfies the balancing property and similar results are presented for 0-1 and S–decomposable fuzzy measures where S is a t–conorm. Moreover, we introduce the concept of distance between two additive fuzzy measures and develop some results related to the distance and the balancing property.

Keywords: fuzzy measures, balancing property, discrete Choquet and Sugeno integrals, aggregation operator, OWA operator, weighted mean.

1 INTRODUCTION

Fuzzy measures and integrals have been introduced by Sugeno in 1974 [17] in order to generalize probability measures. Since that time, they have been studied from different points of view. Particularly, in recent years, there has been an increasing interest in discrete fuzzy integrals, and specially in discrete Choquet and Sugeno integrals, both from a theoretical and an applicational point of view (see e.g. [4, 5, 6, 12, 14, 15, 16]).

On the other hand, another field of increasing interest is related to the aggregation problem. In particular, a lot of authors are interested in the so called aggregation operators. Both problems are closely related because discrete fuzzy integrals can be used as new aggregation tool since they define particular classes of continuous idempotent aggregation operators. It is from the aggregation point of view that we are interested in fuzzy measures and so we use only basic definitions restricted to the discrete case. For more details, we refer to [4, 5, 15, 18].

The purpose of this work is to present a survey of recent results on a subclass of discrete fuzzy measures, i. e., the class of discrete fuzzy measures that satisfy the property called: *balancing property.* The idea of this property is to assign a larger measure to a set of criteria than to any other set with less criteria. As we see along the paper, these results show clearly the situation of fuzzy integrals among aggregation operators, from a theoretical point of view.

2 PRELIMINARIES

In this section, we recall some basic definitions that we shall use throughout this paper. From now on, let us consider a finite universe $X = \{1, \ldots, n\}$.

Definition 1 [17] A fuzzy measure μ on X is a $\mathcal{P}(X) \longrightarrow [0,1]$ mapping that satisfies:

(i) $\mu(\emptyset) = 0$ and $\mu(X) = 1$,

(ii) monotonicity: $A \subset B \Rightarrow \mu(A) \leq \mu(B)$, for all $A, B \subseteq X$.

Definitions on more general spaces usually require algebras and σ–algebras, but this is not necessary in the discrete case. A fuzzy measure on X needs $2^n - 2$ coefficients to be defined, which are the values of μ for all different subsets of X different from X and $\emptyset$.

It is said that a fuzzy measure is

1. a 0–1 *fuzzy measure* if it takes only 0 and 1 as value,
2. *additive* if $\mu(A \cup B) = \mu(A) + \mu(B)$ whenever $A \cap B = \emptyset$,
3. *symmetric* if the following implication holds, for any subsets A and B of X:

$$|A| = |B| \Rightarrow \mu(A) = \mu(B),$$

4. *maxitive* if $\mu(A \cup B) = \mu(A) \vee \mu(B)$,

5. *S-decomposable* measure if $\mu(A \cup B) = S(\mu(A), \mu(B))$ whenever $A \cap B = \emptyset$, with S a continuous t–conorm.

For any fuzzy measure μ, the *dual measure* μ^* is defined by $\mu^*(A) = 1 - \mu(A^c)$ for any $A \subseteq X$. We can quote that if μ is additive then $\mu = \mu^*$, but the converse is not true.

Let us also recall the definitions of discrete Choquet and Sugeno integrals directly as operators from $[0,1]^n$ to $[0,1]$. From now on, given n numbers $x_1, \ldots, x_n$, we will denote by $x_{(1)}, \ldots, x_{(n)}$ the permutation arranged in increasing order: $x_{(1)} \leq \cdots \leq x_{(n)}$.

Definition 2 Given a fuzzy measure μ, the n-dimensional discrete Sugeno integral with respect to μ is the function $S_\mu : [0,1]^n \to [0,1]$ defined by

$$S_\mu(x_1, \ldots, x_n) = \bigvee_{i=1}^{n} (x_{(i)} \wedge \mu(A_{(i)}))$$

where $A_{(i)} = \{(i), (i+1), \ldots, (n)\}$.

Definition 3 Given a fuzzy measure μ, the n-dimensional discrete Choquet integral with respect to μ is the function $C_\mu : [0,1]^n \to [0,1]$ defined by

$$C_\mu(x_1, \ldots, x_n) = \sum_{i=1}^{n} (x_{(i)} - x_{(i-1)})\, \mu(A_{(i)})$$

with the same notations as above, and $x_{(0)} = 0$.

Remark 1 Observe that $A_{(n)} \subset A_{(n-1)} \subset \cdots \subset A_{(2)} \subset A_{(1)} = X$, and $|A_{(i)}| = n - i + 1$, for all $i = 1, \ldots, n$.

The $[0,1]^n \to [0,1]$ mappings C_μ and S_μ that are obtained in this way are always continuous and idempotent aggregation operators. The main results on the relation between fuzzy integrals and aggregation operators are summarized below (see [5, 6, 15]):

- A discrete Choquet integral with respect to an additive fuzzy measure coincides with a weighted arithmetic mean.
- The discrete Choquet integral reduces to the OWA operator for every symmetric fuzzy measure.
- The discrete Sugeno integral corresponds to a weighted median (Kandel and Byatt (1978))

$$S_\mu(x_1, \ldots, x_n) =$$
$$= med(x_1, \ldots, x_n, \mu(A_{(2)}), \ldots, \mu(A_{(n)})).$$

- If μ is a $0-1$ fuzzy measure, then the Choquet and Sugeno integrals are the same aggregation function, which takes the following form:

$$C_\mu(x_1, \ldots, x_n) = S_\mu(x_1, \ldots, x_n) = \bigvee_{A:\mu(A)=1} \left(\wedge_{i \in A} x_{(i)}\right). \quad (1)$$

These are the only continuous aggregation functions verifying $f(x_1, \ldots, x_n) \in \{x_1, \ldots, x_n\}$ [11].

3 BALANCING PROPERTY OF A FUZZY MEASURE

Definition 4 A set function $\mu : \mathcal{P}(X) \to [0,1]$ satisfies the *balancing property* if

$$|A| < |B| \Rightarrow \mu(A) \leq \mu(B) \quad (2)$$

Proposition 1 *Let $\mu : \mathcal{P}(X) \to [0,1]$ be a set function verifying the balancing property.*

(i) If $\mu(\emptyset) = 0$ and $\mu(X) = 1$ then μ is a fuzzy measure and it will be called a BP–fuzzy measure.

(ii) Any symmetric fuzzy measure is a BP–fuzzy measure.

Remark 2 Obviously, there are fuzzy measures such that they are not the BP–fuzzy measures. Note also that a BP–fuzzy measures need not be symmetric.

Proposition 2 *μ is a BP–fuzzy measure if and only if its dual μ^* is also a BP–fuzzy measure.*

3.1 BALANCING PROPERTY OF A 0-1 FUZZY MEASURE

It is possible to determine 0-1 fuzzy measures that satisfy the balancing property and also the aggregation functions defined by this class of measures.

Example 1 For $n = 3$, there are eighteen 0-1 fuzzy measures:

1. Three of them correspond to projections, which are also called Dirac measures. They are the only 0-1 fuzzy measures such that are not BP–fuzzy measures.

2. Three of them correspond to the order statistics, $S_3^1 = \min$, $S_3^2 = med$ and $S_3^3 = \max$. They are the only symmetric 0-1 fuzzy measures.

3. The other 12 measures are not symmetric measures but they are BP–fuzzy measures.

This clasification does not hold for $n > 4$: there are measures that are not BP–fuzzy measures apart from projections.

However, the following result characterizes them, and gives its number:

Proposition 3 *Let μ be a 0-1 BP–fuzzy measure defined on $X = \{1, \ldots, n\}$.*

(i) *There exists s, $0 < s < n$, such that for any $A \subset X$, if $|A| < s$ then $\mu(A) = 0$; if $|A| > s$ then $\mu(A) = 1$ and if $|A| = s$ then $\mu(A) = 0$ or 1.*

(ii) *There are $\sum_{i=1}^{n} 2^{\binom{n}{i}} - n$ of these measures.*

The corresponding discrete integrals can be obtained via (1).

3.2 BALANCING PROPERTY OF AN ADDITIVE FUZZY MEASURE

Here we study the additive measures that verify the balancing property and their corresponding Choquet integrals, the weighted arithmetic means. First, we observe that if μ is additive then it is a discrete probability measure and it is uniquelly determined by the probability vector $(p_1, \ldots, p_n)$, $p_i = \mu(\{(i)\})$. The corresponding Choquet integral coincides with the standard Lebesgue integral and it results to the weighted mean, $C_\mu = \sum_{i=1}^{n} x_i p_i$.

Given a probability vector $p = (p_1, \ldots, p_n)$, let $p^\star = (\alpha_1, \ldots, \alpha_n)$ be a non–decreasing permutation of p.

Proposition 4 *Let μ be an additive fuzzy measure defined on $X = \{x_1, \ldots, x_n\}$.*

(i) *If $n = 2$, then μ is a BP–fuzzy measure.*

(ii) *For $n \geq 3$, μ is a BP–fuzzy measure if and only if*

$$\sum_{k=1}^{r-1} \alpha_{n-k+1} \leq \sum_{k=1}^{r} \alpha_k, \quad r = [\frac{n+1}{2}]. \qquad (3)$$

The above inequality (3) means that the smallest measure of a subset with cardinality r is greater or equal to the largest measure of a subset whit cardinality $r-1$.

For $n = 3$, the above property is $\alpha_3 \leq \alpha_1 + \alpha_2$; for $n = 4$, $\alpha_4 \leq \alpha_1 + \alpha_2$; for $n = 5$, $\alpha_5 + \alpha_4 \leq \alpha_1 + \alpha_2 + \alpha_3$, and so on.

Example 2 We consider a weighted mean with weights $(0.250, 0.335, 0.165, 0.250)$ to give a global evaluation of four different proves. Here, $n = 4$, $r = 2$ and $\alpha_4 = 0.335 \leq \alpha_1 + \alpha_2 = 0.415$, hence condition (2) is satisfied. We can say that they provide a *balancing* between the marks of all proves since the mark of one proof has less significance than the sum of two other marks.

Now, we define a distance between additive measures.

Definition 5 Let μ_1 and μ_2 be two additive fuzzy measures determined by probability vectors $p_1 = (p_1^{(1)}, \ldots, p_n^{(1)})$ and $p_2 = (p_1^{(2)}, \ldots, p_n^{(2)})$. Then the distance between μ_1 and μ_2 is defined by:

$$d(\mu_1, \mu_2) = \sum_{i=1}^{n} |p_i^{(1)} - p_i^{(2)}|. \qquad (4)$$

Note that d is the restriction of the standard L_1–distance in R^n.

Let μ_a be the measure corresponding to the arithmetic mean M_a, i. e., μ_a is an uniformly distributed probability measure defined by the probability vector $p = (\frac{1}{n}, \ldots, \frac{1}{n})$. Then $d(\mu_1, \mu_a) = \sum_{i=1}^{n} |p_i^{(1)} - \frac{1}{n}| = \sum_{i=1}^{n} |\alpha_i - \frac{1}{n}|$, where $(\alpha_1, \ldots, \alpha_n) = p_1^\star$.

Lemma 1 *The distance between any additive measure μ_1 and μ_a is given by*

$$d(\mu_1, \mu_a) = \frac{2q}{n} - 2 \cdot \sum_{i=1}^{q} \alpha_i, \qquad (5)$$

where q is the last index such that $\alpha_q \leq \frac{1}{n}$.

Let us see now how additive BP–fuzzy measures must be.

Proposition 5 *If an additive measure μ is a BP–fuzzy measure then it musts satisfy that:*

$$d(\mu, \mu_a) \leq \frac{2}{n}. \qquad (6)$$

The converse does not hold. Not any measure verifying (6) is a BP–fuzzy measure as we show in the next result:

Lemma 2 *An additive BP–fuzzy measure μ with probability vector p satisfies $d(\mu, \mu_a) = \frac{2}{n}$ if and only if $p^\star = (0, \frac{1}{n-1}, \ldots, \frac{1}{n-1})$.*

There are also additive measures in the interior of $B(\mu_a, \frac{2}{n}) = \{\mu \mid d(\mu_a, \mu) \leq \frac{2}{n}\}$, which are not BP–fuzzy measures. However we get the following result:

Proposition 6 *Let $\mu \in B(\mu_a, \frac{1}{n})$ be an additive measure. Then μ is a BP–fuzzy measure. Moreover, for any $\epsilon > 0$ there exist an additive measure $\mu \in B(\mu_a, \frac{1}{n} + \epsilon)$ which is not a BP–fuzzy measure.*

3.3 BALANCING PROPERTY OF S–DECOMPOSABLE FUZZY MEASURES

In this section, we consider a continuous t–conorm S and analogously to the Section 3.2 it can be shown

Proposition 7 *An S–decomposable fuzzy measure μ is a BP–fuzzy measure if and only if*

1. $\alpha_n = 1 = S(\alpha_1, \alpha_2)$.
2. $\alpha_n < 1$ *and for* $p = \min(r, q-1)$ *it holds*

$$S(\alpha_1, \ldots, \alpha_p) \geq S(\alpha_{n-p+1}, \ldots, \alpha_n), \qquad (7)$$

with $q = \min\{i \in \{2, \ldots, n\} | S(\alpha_1, \ldots, \alpha_i) = 1\}$, $r = [\frac{n+1}{2}]$ *and* $n \geq 3$. *We here consider that* $S(x) = x$ *for all* x *in* $[0, 1]$.

Note that $\alpha_n < 1$ immediatly means that S cannot be a t–conorm such that $S(x, y) = 1$ only if $\max(x, y) = 1$ (i.e., S is either a nilpotent t–conorm, or an ordinal sum with nilpotent summand on the interval $[a, 1], a$ in $[0, 1)$).

Proposition 8 *If* $S(x, y) = 1$ *only when* $\max(x, y) = 1$, *then an S–decomposable fuzzy measure μ is a BP–fuzzy measure if and only if* $\alpha_2 = \alpha_3 = \ldots = \alpha_n = 1$ *and* $\alpha_1 \in [0, 1]$. *Furthermore,*

$$S_\mu(x_1, \ldots, x_n) = \max(x_1, \ldots, x_i \wedge \alpha_1, \ldots, x_n),$$

where $\mu(\{i\}) = \alpha_1$.

In this case, the corresponding Sugeno integral defined by an S–decomposable fuzzy measure is close to the maximum operator.

4 Acknowledgements

The three first authors have been partially supported by CICYT, project TIC96-1393-C06-06 and the fourth author by DGES, project PB96-0191-C02-01.

References

[1] J. Aczél, *On mean values*, Bulletin of the American Math. Society **54** (1948) 392–400.

[2] T. Calvo, B. De Baets and R. Mesiar, *Aggregation operators defined by symmetric k–order additive/maxitive fuzzy measures*, Matware and Soft Computing (1999) (To appear).

[3] G. Choquet, *Theory of capacities*, Ann. Institut Fourier **5** (1953) 131–295.

[4] D. Denneberg, *Non-additive Measure and Integral*, Kluwer Academic, Dordrecht, 1994.

[5] M. Grabisch, H. Nguyen and E. Walker, *Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference*, Kluwer Academic, 1995.

[6] M. Grabisch, *Fuzzy integrals in multicriteria decision making*, Fuzzy Sets and Systems **69** (1995) 279–298.

[7] J. Fodor and M. Roubens, *On meaningfulness of means*, Journal Computacional and Applied Mathematics **64** (1995) 103–115.

[8] A. Kolesárová and M. Komorníková, *Triangular norm-based iterative compensatory operators*, Fuzzy Sets and Systems **104** (1999) 109–119.

[9] A. N. Kolmogoroff, *Sur la notion de la moyenne*, Accad.Naz. Lincei Mem. Cl. Sci. Fis. Mat. Natur. Sez. **12** (1930) 388–391.

[10] G. Mayor and T. Calvo, *On Extended Aggregation Functions*, Proc. Seventh IFSA Congress Vol. **1**, Prague 1997, pp 281–285.

[11] G. Mayor and J. Martín, *Locally internal aggregation functions*, J. Uncertain Fuzziness Knowledge-Based Systems, (Submitted).

[12] R. Mesiar, *Choquet-like integrals*, J. Math. Anal. Appl. **194** (1995) 477–488.

[13] R. Mesiar and M. Komorníková, *Aggregation operators*, Proc. XIth Conference on Appl. Math. PRIM '96 (D. Herceg and K. Surla, eds.), Institute of Mathematics, Novi Sad, 1997, pp. 193–211.

[14] R. Mesiar, *A note to the Choquet integral*, Tatra Mountains, Math, Publ. **12** (1997) 241–245.

[15] T. Murofushi and M. Sugeno, *Some quantities represented by the Choquet integral*, Fuzzy Set and Systems **56** (1993) 229–235.

[16] A. Ralescu and D. Ralescu, *Extensions of fuzzy aggregation*, Fuzzy Sets and Systems **86** (1997) 321–330.

[17] M. Sugeno, *Theory of fuzzy integrals and its applications*, Ph.D. Thesis. Tokyo Institute of Technology, 1974.

[18] Z. Wang and G. Klir, *Fuzzy Measure Theory*, Plenum, 1992.

[19] R. Yager, *On ordered weighted averaging operators in multicriteria decisionmaking*, IEEE Trans. Systems, Man and Cybernetics **18** (1988) 183–190.

ENTROPY OF A CHOQUET CAPACITY*

JEAN-LUC MARICHAL
University of Liège - Faculty of Economics
Sart Tilman - B31 - 4000 Liège, Belgium

MARC ROUBENS
University of Liège - Institute of Mathematics
Sart Tilman - B37 - 4000 Liège, Belgium

1 Introduction

Given a probability distribution $p = (p_1, \ldots, p_n) \in [0,1]^n$, with $\sum_i p_i = 1$, the expression

$$H_n(p) = -\sum_{i=1}^{n} p_i \log_2 p_i,$$

with the convention that $0 \log_2 0 := 0$, is called the *Shannon entropy* of p, see [8, 9]. This value, well-known in information theory, measures in some sense the *uncertainty* which prevailed before an experiment was accomplished, or the *information* expected from the experiment. Note also that it was characterized axiomatically by many authors, see e.g. [1, 3].

Now, consider a Choquet capacity (or fuzzy measure) on $N := \{1, \ldots, n\}$, that is a set function $v : 2^N \to [0,1]$ such that $v(\emptyset) = 0$, $v(N) = 1$, and $v(S) \leq v(T)$ whenever $S \subseteq T$. The following question arises: What is the generalized counterpart of the Shannon entropy for such a capacity?

For particular capacities, such as belief and plausibility measures, some candidates were proposed in evidence theory in the early 1980s, see e.g. [2, 5, 7, 11]. However, it seems that no definition of entropy for a general Choquet capacity was yet proposed in literature.

In this paper we present an entropy-like measure defined for all Choquet capacities. This "entropy" was proposed very recently by Marichal [6] in the framework of aggregation. Although it has yet to be characterized, it satisfies properties considered as requisites for defining an entropy. In particular, it collapses into the Shannon entropy as soon as the capacity is additive.

*To be presented at 1999 Eusflat-Estylf Joint Conference, Palma de Mallorca, September 22-25, 1999

2 Entropy and aggregation operators

Consider the weighted arithmetic mean (WAM) as an aggregation operator:

$$\mathrm{WAM}_\omega(x_1, \ldots, x_n) = \sum_{i=1}^{n} \omega_i x_i$$

with

$$\sum_{i=1}^{n} \omega_i = 1 \quad \text{and} \quad \omega_i \geq 0 \ \forall i = 1, \ldots, n.$$

It is clear that, in such an aggregation process, the use of the information contained in the arguments $x_1, \ldots, x_n$ strongly depends upon the weight vector ω. For example, consider two weighted arithmetic means with weights vectors of the form

$$(1, 0, \ldots, 0) \quad \text{and} \quad (1/n, \ldots, 1/n),$$

respectively. We note that these operators are quite different in the sense that the first one focuses the total weight on only one argument (projection on the first argument) whereas the second one distributes the total weight among all the arguments evenly (arithmetic mean).

In order to capture this idea, one can define a *measure of dispersion* associated to the weight vector of the weighted arithmetic mean WAM_ω as the Shannon entropy of ω:

$$H_n(\omega) = -\sum_{i=1}^{n} \omega_i \log_2 \omega_i.$$

Such a function enables to measure how much of the information in the arguments is really used. In a certain sense the more disperse the ω the more the information contained in the arguments is being used in the aggregation process.

Now, consider the so-called ordered weighted averaging operator (OWA), proposed in 1988 by Yager [12]:

$$\mathrm{OWA}_\omega(x_1, \ldots, x_n) = \sum_{i=1}^{n} \omega_i x_{(i)}$$

with

$$\sum_{i=1}^{n} \omega_i = 1 \quad \text{and} \quad \omega_i \geq 0 \ \forall i = 1, \ldots, n,$$

where $(\cdot)$ indicates a permutation of indices such that $x_{(1)} \leq \cdots \leq x_{(n)}$. For this aggregation operator, the measure of dispersion, which should not depend on a reordering of the arguments, should also be given by the Shannon entropy. In fact, Yager [12] proposed explicitely to use this concept as measure of dispersion for the OWA operators.

It is known that $H_n(\omega)$ is maximum only when ω corresponds to the weight vector of the arithmetic mean, see e.g. [11]:

$$H_n(\omega) = \log_2 n \quad \text{for } \omega = (1/n, \ldots, 1/n),$$

and minimum only when ω is a binary vector:

$$H_n(\omega) = 0 \quad \text{if } \omega_i = 1 \text{ for some } i \in N.$$

Thus, the measure of dispersion can be normalized into

$$\text{disp}(\omega) = \frac{1}{\log_2 n} H_n(\omega) = -\sum_{i=1}^{n} \omega_i \log_n \omega_i,$$

so that it ranges in $[0, 1]$.

3 Entropy of a capacity

Given a Choquet capacity v on N, the *(discrete) Choquet integral* of $x = (x_1, \ldots, x_n)$ w.r.t. v is defined by

$$\mathcal{C}_v(x) = \sum_{i=1}^{n} x_{(i)} \left[v(A_{(i)}) - v(A_{(i+1)}) \right],$$

with the usual convention that $x_{(1)} \leq \cdots \leq x_{(n)}$. Also $A_{(i)} := \{(i), \ldots, (n)\}$, and $A_{(n+1)} = \emptyset$. For more details, see e.g. [4] and the references there.

It is easy to see that the WAM operators correspond to the Choquet integrals w.r.t. additive capacities (i.e., such that $v(S \cup T) = v(S) + v(T)$ whenever $S \cap T = \emptyset$). Moreover, one can show that the OWA operators are exactly those Choquet integrals which are symmetric, that is, independent of any permutation of the arguments. Thus, the Choquet integral is a simultaneous generalization of both WAM and OWA operators.

Starting from these facts, Marichal [6, §6.2.4] proposed to define the entropy of a capacity v as a measure of dispersion for the Choquet integral $\mathcal{C}_v$. This measure should identify with the Shannon entropy when the Choquet integral is either a WAM or an OWA.

On the one hand, comparing

$$\text{OWA}_\omega(x) = \sum_{i=1}^{n} x_{(i)} \, \omega_i$$

and

$$\mathcal{C}_v(x) = \sum_{i=1}^{n} x_{(i)} \left[v(A_{(i)}) - v(A_{(i+1)}) \right]$$

suggests to propose as measure of dispersion for $\mathcal{C}_v$ a sum over $i \in N$ of an average value of

$$[v(T \cup \{i\}) - v(T)] \log_n [v(T \cup \{i\}) - v(T)], \ T \subseteq N \backslash \{i\},$$

that is an expression of the form

$$\begin{aligned} \text{disp}(v) \quad = \quad & -\sum_{i=1}^{n} \sum_{T \subseteq N \backslash \{i\}} p_{|T|} \left[v(T \cup \{i\}) - v(T) \right] \\ & \times \log_n [v(T \cup \{i\}) - v(T)], \end{aligned}$$

where the coefficients $p_{|T|}$ are non-negative and such that $\sum_{T \subseteq N \backslash \{i\}} p_{|T|} = 1$.

On the other hand, imposing the condition

$$\mathcal{C}_v = \text{OWA}_\omega \quad \Rightarrow \quad \text{disp}(v) = \text{disp}(\omega)$$

determines uniquely the coefficients $p_{|T|}$, so that the definition proposed is the following.

Definition 1 *The entropy of a Choquet capacity v on N is defined by*

$$\begin{aligned} \text{disp}(v) := & -\sum_{i=1}^{n} \sum_{T \subseteq N \backslash \{i\}} \frac{(n-t-1)!\, t!}{n!} \\ & \times [v(T \cup \{i\}) - v(T)] \log_n [v(T \cup \{i\}) - v(T)]. \end{aligned}$$

When the Choquet integral $\mathcal{C}_v$ is used as an aggregation operator, this entropy can be interpreted as the degree to which one uses all the information contained in the arguments $x = (x_1, \ldots, x_n)$ when calculating the aggregated value $\mathcal{C}_v(x)$.

Interestingly enough, its expression is very similar to that of the Shapley value of elements in N, which is a fundamental concept in game theory [10] expressing a power index:

$$\phi_i(v) = \sum_{T \subseteq N \backslash \{i\}} \frac{(n-t-1)!\, t!}{n!} \left[v(T \cup \{i\}) - v(T) \right],$$

$i \in N$.

Notice also that this new definition has yet to be axiomatically characterized. However, to justify its use, one can show that it fulfils several properties required for an entropy [6], namely:

- $\mathrm{disp}(v)$ is *continuous* w.r.t. v.
- disp is *symmetric*, that is

$$\mathrm{disp}(\pi v) = \mathrm{disp}(v)$$

for any permutation π of N. Here, πv is the capacity on N defined by $\pi v(\pi(S)) = v(S)$ for all $S \subseteq N$, where $\pi(S) = \{\pi(i) \mid i \in S\}$.

- We have

$$0 \leq \mathrm{disp}(v) \leq 1.$$

Moreover, $\mathrm{disp}(v)$ is maximum ($= 1$) only when $\mathcal{C}_v$ is the arithmetic mean, and minimum ($= 0$) only when v is a binary-valued capacity: $v(S) \in \{0, 1\}$ for all $S \subseteq N$. Note that this latter case occurs if and only if $\mathcal{C}_v(x) \in \{x_1, \ldots, x_n\}$ (only one piece of information is used in the aggregation).

- We have

$$\mathcal{C}_v = \mathrm{WAM}_\omega \text{ or } \mathrm{OWA}_\omega$$
$$\Downarrow$$
$$\mathrm{disp}(v) = \mathrm{disp}(\omega).$$

- Let $k \in N$ be a *null* element for v, that is, $v(T \cup \{k\}) = v(T)$ for all $T \subseteq N \setminus \{k\}$. Then

$$\mathrm{disp}(v) = \mathrm{disp}(v^{N \setminus \{k\}})$$

where $v^{N \setminus \{k\}}$ is the restriction of v to $N \setminus \{k\}$.

References

[1] J. Aczél and Z. Daróczy, *On measures of information and their characterizations*, (Academic Press, New York–San Francisco–London, 1975).

[2] D. Dubois and A. Ramer, Extremal properties of belief measures in the theory of evidence, *Int. J. Uncertainty, Fuzziness and Knowledge-Based Systems* **1** (1993) 57–68.

[3] B. Ebanks, P. Sahoo, and W. Sander, *Characterizations of information measures*, (World Scientific, Singapore, 1997).

[4] M. Grabisch, The application of fuzzy integrals in multicriteria decision making, *European Journal of Operational Research*, **89** (1996) 445-456.

[5] G.J. Klir and A. Ramer, Uncertainty in the Dempster-Shafer theory: a critical re-examination, *Int. J. General Systems* **18** (1990) 155–166.

[6] J.-L. Marichal, *Aggregation operators for multicriteria decision aid*, Ph.D. thesis, Institute of Mathematics, University of Liège, Liège, Belgium, 1998.

[7] A. Ramer, Uniqueness of information measure in the theory of evidence, *Fuzzy Sets and Systems* **24** (1987) 183–196.

[8] C.E. Shannon, A mathematical theory of communication, *Bell System Tech. J.* **27** (1948) 379–423, 623–656.

[9] C.E. Shannon and W. Weaver, A mathematical theory of communication, (University of Illinois Press, Urbana, 1949).

[10] L.S. Shapley, A value for n-person games. In: H.W. Kuhn and A.W. Tucker (eds.), *Contributions to the Theory of Games*, Vol. II, Annals of Mathematics Studies, 28, (Princeton University Press, Princeton, NJ, 1953), 307–317.

[11] R.R. Yager, Entropy and specificity in a mathematical theory of evidence, *Int. J. Gen. Systems* **9** (1983) 249–260.

[12] R.R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, *IEEE Trans. on Systems, Man and Cybernetics* **18** (1988) 183–190.

GENERATING HASSE TREES OF FUZZY PREORDER CLOSURES: AN ALGORITHMIC APPROACH

Helga Naessens, Bernard De Baets, Hans De Meyer
Department of Applied Mathematics and Computer Science
University of Ghent
Krijgslaan 281 (S9), B-9000 Gent, Belgium
{Helga.Naessens,Bernard.DeBaets,Hans.DeMeyer}@rug.ac.be

Summary

A top-down algorithm for generating the Hasse tree of the fuzzy preorder closure of an arbitrary reflexive binary fuzzy relation is presented.

Keywords: fuzzy preorder, transitive closure, Hasse tree.

1 INTRODUCTION

Mathematical theories are pervaded with the use of partial orders, and more general, preorders (also called quasi-orders), i.e. reflexive and transitive binary relations. Preorders also play an important role in applications, for instance as preference relations in multicriteria decision aid techniques. The observation that classical relations do not allow to express partial or graded relationships has led to the introduction of fuzzy relations. In preference modelling, for instance, fuzzy relations are used to express graded preferences, indifferences and incomparabilities [4, 10].

In this paper, we will consider this decision-theoretic framework in which fuzzy preorders, i.e. reflexive and transitive binary fuzzy relations, are used as large preference relations. An important issue is that of the transitivity of the large preference relation. Transitivity is not always ensured (or even not desired) but can always be enforced by considering the transitive closure, which in this case is called fuzzy preorder closure.

Fuzzy preorders admit a nice graphical representation by means of a Hasse tree, i.e. a partition tree with Hasse diagrams drawn on top of it. Such a Hasse tree forms a suitable basis for the interpretation of preferential relationships in the set of alternatives. For more details about the exploitation of these Hasse trees, we refer to [3, 4]. Applications can be found in various disciplines such as cognitive psychology [1], finance [6], nuclear engineering [9] and geography [5].

In this paper, we focus on the problem of generating the Hasse tree. We present a top-down algorithm that allows to produce the Hasse tree level by level for decreasing degrees of confidence. Another advantage of the algorithm is its applicability for any binary reflexive fuzzy relation, in the sense that it produces at the same time the fuzzy preorder closure.

2 BASIC NOTIONS

We first recall some definitions that will be used throughout. A binary fuzzy relation R on a universe X is called a fuzzy preorder if it is reflexive and transitive, i.e. for any (x, y, z) in X^3 it holds that $R(x, x) = 1$ and

$$\min(R(x, y), R(y, z)) \leq R(x, z).$$

For α in $[0, 1]$, the α-cut R_α of a binary fuzzy relation R on X is the crisp binary relation on X defined by

$$(x, y) \in R_\alpha \Leftrightarrow R(x, y) \geq \alpha\,.$$

An important theorem states that R is a fuzzy preorder if and only if every α-cut R_α is a preorder [11]. This implies that for each α, R_α induces an equivalence relation E_α on X and a partial order $\leq_\alpha$ on the quotient set X/E_α. Formally,

$$(x_i, x_j) \in E_\alpha \Leftrightarrow (x_i, x_j) \in R_\alpha \wedge (x_j, x_i) \in R_\alpha\,,$$

and

$$[x_i]_\alpha \leq_\alpha [x_j]_\alpha \Leftrightarrow (x_i, x_j) \in R_\alpha \wedge (x_j, x_i) \notin R_\alpha\,,$$

where $[x_i]_\alpha$ and $[x_j]_\alpha$ are the equivalence classes of x_i and x_j w.r.t. E_α. It is clear that with decreasing α, the equivalence classes tend to merge together, or, in decision-theoretic terms: *"incomparability disappears at the cost of increasing indifference"* [3].

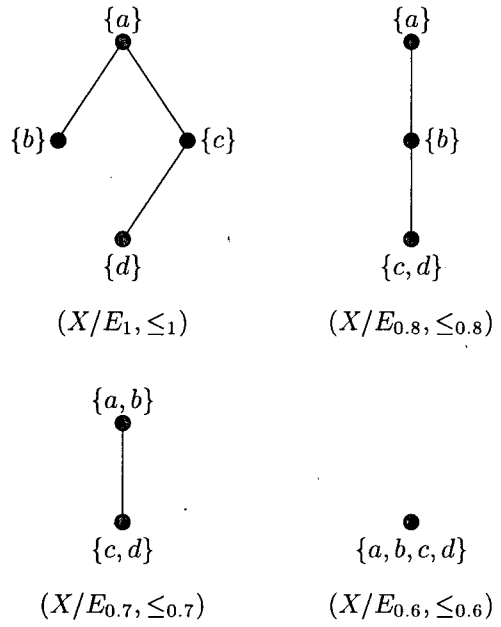


Figure 1: Hasse diagrams

As an illustration, consider the fuzzy preorder R on $X = \{a, b, c, d\}$ given by:

R	a	b	c	d
a	1	0.7	0.6	0.6
b	1	1	0.6	0.6
c	1	0.8	1	0.8
d	1	0.8	1	1

(1)

The Hasse diagrams of the four corresponding partial orders are depicted in Figure 1. These diagrams can be summarized in the Hasse tree drawn in Figure 2. In this tree, the broken lines in each level (identified by the α-cut R_α) represent the arcs of the Hasse diagram of $(X/E_\alpha, \leq_\alpha)$, while the nodes represent the equivalence classes of E_α.

For a non-transitive reflexive binary fuzzy relation R, the Hasse tree is drawn from its fuzzy preorder closure $\hat{R}$, which is defined as the smallest fuzzy preorder containing R, i.e. a fuzzy preorder $\hat{R}$ such that any fuzzy preorder E satisfying $R \subseteq E \subseteq \hat{R}$ coincides with $\hat{R}$ [2].

3 GENERATING HASSE TREES

In this section, we present a novel algorithm (Algorithm 1) for generating the Hasse tree of the fuzzy preorder closure of an arbitrary reflexive binary fuzzy relation R. A reflexive binary fuzzy relation R on X can be diagrammed by a *weighted* reflexive directed graph, whereby the vertices represent the elements of X and whereby the weights of the arcs are the corresponding values of R. Then, a reflexive binary fuzzy relation is

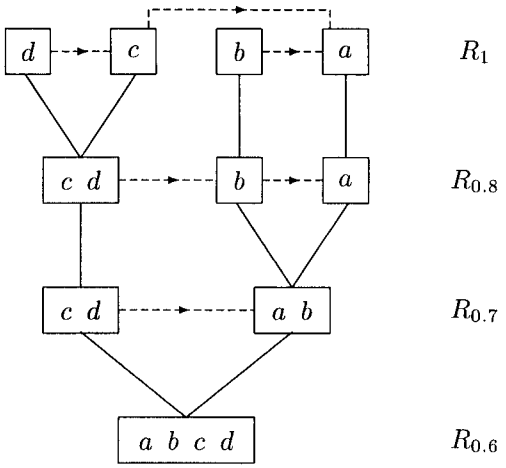


Figure 2: Hasse tree for R

a fuzzy preorder if and only if all the triangular subgraphs of its corresponding graph are transitive.

Algorithm 1 is inspired on a new algorithm we have developed for the computation of the T-transitive closure of a proximity relation, whereby T can be any triangular norm [7, 8]. It acts on a universe X with cardinality n, the elements of which are simply referred to as 1, 2, ..., n.

On execution of Algorithm 1 the arcs of the directed graph are visited in descending order of their actual weights (double repeat loop: an outer loop running over the different weights in the graph and an inner loop running over all the arcs with the same weight) and used as pivot, whereby it is taken into account that weights can change during execution of the algorithm. At the stage whereby arc (i, j) is the pivot, all non-transitive triangles containing arc (i, j) (for loop) and in which w_{ij} is the maximum weight (if tests) are made transitive by raising the smallest weight in the triangle.

Next, we add arc (i, j) to the set NewArcs. Then the algorithm checks if the class containing element i and the class containing element j merge. This is the case when arc (j, i) has previously been used as pivot. If the two classes merge, we add all the elements of the class containing j to the class containing i and we update the sets NewArcs and OldArcs. This means that we add arc (i, k), for all $k \in S \setminus \{i, j\}$, to NewArcs if OldArcs or NewArcs contains (j, k), but none of them contains (i, k). Analogously, we add arc (k, i) to NewArcs, if required. Then we remove all the arcs through vertex j from OldArcs and NewArcs to prevent them from being selected as pivot later on. Next, vertex j is removed from the set S, which will contain, upon termination of the inner loop, the representatives of the classes at the corresponding level.

After having performed these actions for pivots with the same weight (inner repeat loop), Algorithm 1 removes redundant arcs from the sets NewArcs and OldArcs, which requires two steps. In both steps, we make use of the fact that an arc (i,j) may be removed if there exists an indirect path from i to j in OldArcs $\cup$ NewArcs, which can be shown to be equivalent with $w_{ij} \geq \alpha$.

Algorithm 1 Generate the Hasse tree of the fuzzy preorder closure $\hat{R}$ of a fuzzy relation R

Input :
Vertex Set $= \{1, 2, \ldots, n\}$;
Weight Set $= \{w_{ij} = R(i,j) \mid \forall i, j \in$ Vertex Set$\}$;
Output :
Description of Hasse tree of $\hat{R}$;

begin
  $V := \{(i,j) \mid \forall i \neq j \in$ Vertex Set$\}$;
  $S := \{1, 2, \ldots, n\}$;
  for $k := 1$ **to** n **do** class$[k] := \{k\}$;
  OldArcs $:= \emptyset$;
  repeat
    Select an arc $(i,j) \in V$ such that
      $w_{ij} \geq w_{lm}\ \forall (l,m) \in V$;
    level $:= w_{ij}$;
    NewArcs $:= \emptyset$;
    repeat
      for $k := 1$ **to** n **do**
        if $k \in S$ **and** $w_{jk} \leq w_{ij}$ **and** $w_{ik} < w_{jk}$ **then**
          $w_{ik} := w_{jk}$;
        if $k \in S$ **and** $w_{ki} \leq w_{ij}$ **and** $w_{kj} < w_{ki}$ **then**
          $w_{kj} := w_{ki}$
      endfor;
      NewArcs $:=$ NewArcs $\cup \{(i,j)\}$;
      $V := V \setminus \{(i,j)\}$;
      if class$[i]$ and class$[j]$ merge **then**
        class$[i]$:= class$[i] \cup$ class$[j]$;
        class$[j] := \emptyset$;
        Update NewArcs and OldArcs;
        $V := V \setminus \{(j,1), \ldots, (j,n), (1,j), \ldots, (n,j)\}$;
        $S := S \setminus \{j\}$
      endif;
      Select an arc $(i,j) \in V$ such that
        $w_{ij} \geq w_{lm}\ \forall (l,m) \in V$
    until $w_{ij} \neq$ level;
    Remove redundant arcs from NewArcs;
    Remove redundant arcs from OldArcs;
    OldArcs := OldArcs $\cup$ NewArcs;
    Write level, classes and arcs of OldArcs
  until $V = \emptyset$
end

The first step (Procedure 1) removes redundant arcs from the set NewArcs, for which it investigates in pairs the arcs of NewArcs. Suppose that arcs (i,j) and (k,l) are selected. Then, arc (i,j) may be removed if $w_{ik} \geq \alpha$ and $w_{lj} \geq \alpha$, since in that case there exists a path from i to j, containing arc (k,l).

Procedure 1 Remove redundant arcs from NewArcs

begin
  $X :=$ NewArcs;
  repeat
    Select an arc $(i,j) \in X$;
    $Y :=$ NewArcs $\setminus \{(i,j)\}$;
    repeat
      Select an arc $(k,l) \in Y$;
      if $w_{ik} \geq \alpha$ **and** $w_{lj} \geq \alpha$ **then**
        NewArcs := NewArcs $\setminus \{(i,j)\}$;
      $Y := Y \setminus \{(k,l)\}$
    until $Y = \emptyset$;
    $X := X \setminus \{(i,j)\}$
  until $X = \emptyset$
end

The second step (Procedure 2) removes redundant arcs from the set OldArcs, for which it investigates in pairs the arcs of OldArcs and NewArcs. Suppose that arcs (i,j) from OldArcs and (k,l) from NewArcs are selected. Then, arc (i,j) may be removed if $w_{ik} \geq \alpha$ and $w_{lj} \geq \alpha$, since in that case there again exists a path from i to j, containing arc (k,l).

Procedure 2 Remove redundant arcs from OldArcs

begin
  $X :=$ OldArcs;
  repeat
    Select an arc $(i,j) \in X$;
    $Y :=$ NewArcs;
    repeat
      Select an arc $(k,l) \in Y$;
      if $w_{ik} \geq \alpha$ **and** $w_{lj} \geq \alpha$ **then**
        OldArcs := OldArcs $\setminus \{(i,j)\}$;
      $Y := Y \setminus \{(k,l)\}$
    until $Y = \emptyset$;
    $X := X \setminus \{(i,j)\}$
  until $X = \emptyset$
end

Next, Algorithm 1 adds the arcs of NewArcs to the set OldArcs. As mentioned above, the elements of S are the representatives of the classes at the corresponding level, and the graph $(S,$ OldArcs$)$ represents the Hasse diagram of this level. Finally, we write out the level, the classes and the arcs of OldArcs.

4 EXAMPLE

To illustrate our algorithm, let us consider the fuzzy relation S given by:

S	a	b	c	d
a	1	0.7	0.4	0.5
b	1	1	0.6	0.4
c	1	0.8	1	0.8
d	1	0.7	1	1

Obviously, S is not transitive since for example $\min(S(a,b), S(b,c)) > S(a,c)$. The fuzzy preorder closure of S is nothing else but the fuzzy relation R (1). If we run Algorithm 1 on the fuzzy relation S, the following output is generated:

```
Level 1.0 :
Class 1 : a
Class 2 : b
Class 3 : c
Class 4 : d
Arc from Class 2 to Class 1
Arc from Class 3 to Class 1
Arc from Class 4 to Class 3

Level 0.8 :
Class 1 : a
Class 2 : b
Class 3 : c d
Arc from Class 2 to Class 1
Arc from Class 3 to Class 2

Level 0.7 :
Class 1 : a b
Class 2 : c d
Arc from Class 2 to Class 1

Level 0.6 :
Class 1 : a b c d
```

As expected, this is a description of the Hasse tree shown in Figure 2.

Acknowledgements

H. De Meyer is a Research Director and B. De Baets is a Post-Doctoral Fellow of the Fund for Scientific Research - Flanders.

References

[1] W. Bandler and L.J. Kohout, *Fuzzy relational products as a tool for analysis and synthesis of the behaviour of complex natural and artificial systems*, Fuzzy Sets: Theory and Application to Policy Analysis and Information Systems (S. Wang and P. Chang, eds.), Plenum Press, New York and London, 1980, pp. 341–367.

[2] W. Bandler and L.J. Kohout, *Special properties, closures and interiors of crisp and fuzzy relations*, Fuzzy Sets and Systems **26** (1988) 317–331.

[3] B. De Baets, *Ordering alternatives in MCDM problems*, Fuzzy Logic and Intelligent Technologies in Nuclear Science (D. Ruan et al., eds.), Proc. First International FLINS Workshop (Mol, Belgium), World Scientific Publishing, Singapore, 1994, pp. 48–53.

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

[5] R. Groenemans, E. Kerre and E. Van Ranst, *Fuzzy relational calculus in land evaluation*, GEODERMA **77** (1997) 283–298.

[6] E. Haven, *The fuzzy multicriteria analysis method: an application on NPV ranking*, Int. J. Intell. Sys. Acc. Fin. Mgmt **7** (1998) 243–252.

[7] H. Naessens, H. De Meyer and B. De Baets, *Novel algorithms for the computation of T-transitive closures and openings of proximity relations*, Proc. EUROFUSE-SIC '99 (Budapest, Hungary) (B. De Baets, J. Fodor, L. Kóczy, eds.), 1999, pp. 200-203.

[8] H. Naessens, H. De Meyer and B. De Baets, *Algorithms for the computation of the T-transitive closure of a fuzzy relation*, in preparation.

[9] B. Van de Walle, B. De Baets and E. Kerre, *Fuzzy multi-criteria analysis of cutting techniques in a nuclear reactor dismantling project*, Fuzzy Sets and Systems, Special Issue "Nuclear Engineering" (D. Ruan, ed.), **74** (1995) 115–116.

[10] B. Van de Walle, B. De Baets and E. Kerre, *Characterizable fuzzy preference structures*, Ann. Oper. Res., Special Issue "Preference Modelling" (D. Bouyssou and Ph. Vincke, eds.), **80** (1998) 105-136.

[11] L. Zadeh, *Similarity relations and fuzzy orderings*, Inform. Sci. **3** (1971) 177–200.

ALGORITHMS TO EXTEND CRISP FUNCTIONS AND THEIR INVERSE FUNCTIONS TO FUZZY NUMBERS

O.G. Duarte
Department of Electrical Engineering
Universidad Nacional de Colombia
Ciudad Universitaria Ed 453 Of 202
Bogotá, Colombia
ogduarte@ingenieria.ingsala.unal.edu.co

M. Delgado
Department of Computer Sciences
and A.I. - E.T.T.S.I. Informática,
Universidad de Granada,
Avda de Andalucía 38
C.P. 18071 Granada, Spain
mdelgado@decsai.ugr.es

I.Requena
Department of Computer Sciences
and A.I. - E.T.T.S.I. Informática,
Universidad de Granada,
Avda de Andalucía 38
C.P. 18071 Granada, Spain
irequena@decsai.ugr.es

Abstract

In this paper we show an algorithm to extend a kind ofcontinuous crisp functions to fuzzy numbers using the extension principle. Then we present two different solutions to the problem of extending inverse functions which we have called possible extension and necessary extension. Finally, using these solutions we have generated a family of intermediate extensions that let us define a parameter to measure the existence of the extended inverse functions.

Keywords: Fuzzy numbers, fuzzy relations, possible extension, necessary extension.

1 INTRODUCTION

Although fuzzy numbers can successfully represent the uncertainty of a numerical variable in a mathematical model, there are some difficulties representing the functions that relate these variables, mainly because with the most usual definition (used in this paper and shown later), the fuzzy numbers do not have a group algebraic structure.

Another difficulty is that mathematical models are generally designed to deal with non- fuzzy (crisp) numerical variables and therefore the functions in the models operate on crisp numbers. If we attempt to use fuzzy numbers to represent numerical variables, it is necessary to extend the crisp functions included in the model to fuzzy numbers. The usual way to extend a function is to use the Extension Principle postulated by Zadeh [4]; although this is generally an effective and useful strategy, in some cases it is not the best way to deal with inverse functions, because there are special difficulties associated with inverse functions; this has been studied in depth by Yager [3]. Bouchon- Meunier et al [1] have demonstrated that if a crisp function is invertible then fuzzy numbers are the only fuzzy sets that maintain the invertibility of the function.

Furthermore, as Giachetti [2] notes, when dealing with functions of more than one argument we can have diverse definitions of "invertibility" according to the application of the inverse function; we explain it in section 4.

The main aim of this paper is to present some algorithms that allow functions which relate crisp numerical variables in a mathematical model to be used when these variables are represented by fuzzy numbers, without the need to modify the original crisp functions. The algorithms related to inverse functions ensure the existence of a solution, but sometimes it is necessary to modify the problem conditions. furthermore, we develop a measure of the existence of the extended inverse function. The functions included in this paper are multiple argument functions, monotonically increasing in some of the arguments and monotonically decreasing in the others.

2 BASIC DEFINITIONS AND NOTATION

2.1 DEFINITION 1: FUZZY NUMBERS

We use in our algorithms the following representation of a Fuzzy Number: Let A be a number with membership function $A(x)$; the α-cuts of A are:

$$A_\alpha = \{x|A(x) > \alpha\} = [L_A(\alpha), R_A(\alpha)]\alpha \in (0,1] \quad (1)$$

A is a fuzzy number if and only if $L_A(\alpha)$ and $R_A(\alpha)$ satisfy that $L_A(\alpha)$ must be monotonically increasing and right continuous, $R_A(\alpha)$ must be monotonically decreasing and left continuous, and $L_A(1) \leq R_A(1)$

A may also be represented using the function $D_A(\alpha, d) : (0,1] \times \{-1,1\} \rightarrow R$ as follows:

$$D_A(\alpha, d) = \begin{cases} L_A(\alpha) & \text{if } d = 1 \\ R_A(\alpha) & \text{if } d = -1 \end{cases}$$

The α-cuts of A may now be written:

$$A_\alpha = [D_A(\alpha, 1), D_A(\alpha, -1)]$$

Note that the representation of a fuzzy number using the function $D_A(\alpha, d)$ is conceptually the same as if we use α-cuts. In this paper we will use the function $D_A(\alpha, d)$ so that the algorithms are more compact.

In our algorithms we will suppose that we want to know a discrete number of α-cuts, associated with an ordered set $Alfa = \{\alpha_1, \alpha_2, ..., \alpha_p\}$ where $\alpha_i < \alpha_{i+1}$ and $\alpha_1 \geq 0$.

2.2 DEFINITION 2

In the rest of the paper we assume that $y = f(X) : U_1 \times U_2 \times \ldots \times U_n \rightarrow V$ is a continuous function, monotonically increasing with some of the n variables and monotonically decreasing with the others ($U_i, V \subseteq R$). We also assume that $x_k = f_k^{-1}(y, X_k)$ $X_k = \{x_1, x_2, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n\}$ is the function that calculates the value of x_k when we know the output y and the other input variables x_i; furthermore, we assume that $f(X)$ and $f_k^{-1}(y, X_k)$ are well defined in all the possible combinations of their respective arguments. We define the function $d_f(i) : I \rightarrow \{1, -1\}$ (I is the index set) as follows:

$$d_f(i) = \begin{cases} 1 & \text{if } f(X) \text{ increases with } x_i \\ -1 & \text{if } f(X) \text{ decreases with } x_i \end{cases}$$

2.3 DEFINITION 3

Let $A_1, A_2, \ldots, A_n$ be n fuzzy numbers defined on $U_1, U_2, \ldots U_n$ respectively using the functions $D_{Ai}(\alpha, d)$ (definition 1) with A_i representing the value of the argument x_i of $f(X)$.We define

$$\begin{aligned} D_X(\alpha, d_f) &= \{D_{Ai}(\alpha, d_f(i)); i = 1, \ldots, n\} \quad (2) \\ D_X(\alpha, -d_f) &= \{D_{Ai}(\alpha, -d_f(i)); i = 1, \ldots, n\} \end{aligned}$$

Each of the elements $D_{Ai}(\alpha, d_f(i))$ and $D_{Ai}(\alpha, -d_f(i))$ belong to the α-cut of A_i. Furthermore, they are the lowest or highest value of the α-cut of A_i depending on the increasing-decreasing condition of the variable x_i. So, $D_X(\alpha, d_f)$ is the set of values of X belonging to the respective α-cuts of $X_1, X_2, \ldots, X_n$ that gives y the lowest value. Moreover, $D_X(\alpha, -d_f)$ is the set of values of X belonging to the respective α-cuts of$X_1, X_2, \ldots, X_n$ that gives y the highest value. Note that $D_{Ai}(\alpha, d_f(i))$ is an increasing function of α, whereas $D_{Ai}(\alpha, -d_f(i))$ is a decreasing one. We also define

$$D_X(\alpha, d_f(k)) = \{D_{Ai}(\alpha, d_f(k) d_f(i)); i = 1, \ldots, n; i \neq k\}$$

$$D_X(\alpha, -d_f(k)) = \{D_{Ai}(\alpha, -d_f(k) d_f(i)); i = 1, \ldots, n; i \neq k\}$$

3 EXTENDING CRISP FUNCTIONS TO FUZZY NUMBERS

Let $y = f(X)$ be a crisp function with the characteristics indicated in definition 2. Suppose we want to extend it to fuzzy numbers; that is, we want to find a function that allows us to calculate y (or more exactly, a fuzzy number representing y) when we represent the n arguments of $f(X)$ using the fuzzy numbers $A_1, A_2, \ldots, A_n$. Due to the continuity of $f(X)$, we can extend the function $y = f(X)$ to fuzzy numbers using the Extension Principle with the α-cuts as follows:

$$y_\alpha = f(A_{1\alpha}, A_{2\alpha}, \ldots, A_{n\alpha})$$

The above expression implies that an α-cut of y is the set of all the values obtained when all the possible values of x_i belonging to the respective α- cuts of the input variables are used as arguments of $f(X)$. As $f(X)$ is monotonically increasing or decreasing with all its arguments, and since the α- cuts of the input variables are closed intervals, then y_α is also a closed interval whose extremes can be calculated using (3), according to definition 3.

$$y_\alpha = [f(D_X(\alpha, d_f)), f(D_X(\alpha, -d_f))] \quad (3)$$

As stated in definition 3, $L_y(\alpha)$ is increasing and $R_y(\alpha)$ is decreasing; furthermore $L_{y1} \leq R_{y1}$. Due to the continuity of $f(x)$ then the conditions in (1) are satisfied, and y_α represents a fuzzy number. If we have a discrete representation as proposed in 2.1, then the algorithm 1 is used to extend $y = f(X)$ to fuzzy numbers.

Algorithm 1

1	For α_j, $j = 1, 2, \ldots, p$ calculate y_{α_j} using (3)

4 EXTENDING CRISP INVERSE FUNCTIONS TO FUZZY NUMBERS

Taking the same conditions as in section 3, suppose that we now want to extend the crisp inverse function $x_k = f_k^{-1}(y, X_k)$, with $X_k = \{x_1, x_2, \ldots, x_{k-1}, x_{k+1}, \ldots, x_n\}$ which calculates the value of the input variable x_k when we know the output y and the other input variables x_i. The solution to this problem may have at least two different approaches when each variable is represented by a fuzzy number:

- We want to know which values the variable x_k could take if we know the values of y, X_k . In

this paper, this approach will be called the *possible extension* of the inverse function, and will be denoted by x_k^{pos}.

- We want to know which values the variable x_k must take in order to ensure that the output variable y takes certain predetermined values, when we know the values of the other input variables x_K. In this paper, this approach will be called the *necessary extension* of the inverse function, and will be denoted by x_k^{nec}.

4.1 POSSIBLE EXTENSION

The procedure used to obtain the possible extension consists in considering the crisp inverse function $x_k = f_k^{-1}(y, X_k)$ as a direct function, and extending it using the procedure described in section 3. If we take into account how x_k varies with regard to the other variables we can write an equation (4) to calculate $x_k^{pos}(\alpha)$.

$$\begin{aligned} x_k^{pos}(\alpha) &= [Lx_k^{pos}(\alpha), Rx_k^{pos}(\alpha)] \\ Lx_k^{pos}(\alpha) &= f_k^{-1}(D_y(\alpha, d_f(k)), D_{Xk}(\alpha, -d_{fk})) \\ Rx_k^{pos}(\alpha) &= f_k^{-1}(D_y(\alpha, -d_f(k)), D_{Xk}(\alpha, d_{fk})) \end{aligned} \quad (4)$$

On account of having extended the crisp function f_k^{-1} to fuzzy numbers using the procedure of section 3, $x_k^{pos}(\alpha)$ represents a fuzzy number; $D_y(\alpha, d_f(k))$, $D_{Xk}(\alpha, d_{fk})$ and $D_{Xk}(\alpha, -d_{fk})$ allow us to write in shorthand if x_k decreases or increases when the other variables increase.

If we have a discrete representation as proposed in 2.1, then the algorithm 2 is used to make a possible extension of $x_k = f_k^{-1}(y, X_k)$ to fuzzy numbers:

Algorithm 2

1	For α_j, $j = 1, 2, \ldots, p$ calculate $x_k^{pos}(\alpha_j)$ using (4)

4.2 NECESSARY EXTENSION

The procedure used to obtain the necessary extension consists in calculating the extremes of every α-cut of x_k using $x_k = f_k^{-1}(y, X_k)$ with the most restrictive extremes of the α-cuts of y and X_k. The expression used to calculate $x_k^{nec}(\alpha)$ is (5).

$$\begin{aligned} x_k^{nec}(\alpha) &= [Lx_k^{nec}(\alpha), Rx_k^{nec}(\alpha)] \\ Lx_k^{nec}(\alpha) &= f_k^{-1}(D_y(\alpha, d_f(k)), D_{Xk}(\alpha, d_{fk})) \\ Rx_k^{nec}(\alpha) &= f_k^{-1}(D_y(\alpha, -d_f(k)), D_{Xk}(\alpha, -d_{fk})) \end{aligned} \quad (5)$$

However, we cannot ensure that $Lx_k^{nec}(\alpha)$ will always be monotonically increasing, nor that $Rx_k^{nec}(\alpha)$ will always be monotonically decreasing, nor that $Lx_k^{nec}(\alpha) \leq Rx_k^{nec}(\alpha)$ for every α, and as a result, (5) will not generally represent a fuzzy number. Only for certain fuzzy numbers Y will we have a fuzzy number represented by $x_k^{nec}(\alpha)$ in (5). The reason for this is explained in the following example:

Algorithm 3

1	For $\alpha_p = 1$ - calculate $Lx_k^{nec}(\alpha_p)$, $Rx_k^{nec}(\alpha_p)$ using (5) - if $Lx_k^{nec}(\alpha_p) > Rx_k^{nec}(\alpha_p)$ then widen the α-cut $Y\alpha_p = [L_y(\alpha_p), R_y(\alpha)_p)]$ in the left and right sides until we get a new α-cut $Y'\alpha_p = [L_y'(\alpha_p), R_y'(\alpha)_p)]$ such that $Lx_k^{nec}(\alpha_p) \leq Rx_k^{nec}(\alpha_p)$
2	For α_j, $j = p-1, p-2, \ldots, 1$ - if $L_y(\alpha_j) > L_y'(\alpha_{j+1})$ let $L_y(\alpha_j) = L_y'(\alpha_{j+1})$ - if $R_y(\alpha_j) < R_y'(\alpha_{j+1})$ let $R_y(\alpha_j) = R_y'(\alpha_{j+1})$
3	For α_j, $j = p-1, p-2, \ldots, 1$ - calculate $Lx_k^{nec}(\alpha_j)$, using (5) - if $Lx_k^{nec}(\alpha_j) > Lx_k^{nec}(\alpha_{j+1})$ then calculate y_α^* as follows: $y_\alpha^* = f(D_X(\alpha_j^*, d_f))$ $D_X(\alpha_j^*, d_f) = \{D_X(\alpha_{ji}^*, d_f(i) d_f(k)$ $\mid i = 1, 2, \ldots, n\}$ $\alpha_{ji}^* = \begin{cases} \alpha_j & \text{if } i \neq k \\ \alpha_{j+1} & \text{if } i = k \end{cases}$ - if $d_f(k) = 1$ then widen the α-cut $Y\alpha_j = [L_y(\alpha_j), R_y(\alpha)_j)]$ in the left side transforming it into the new α-cut $Y'\alpha_j = [y_\alpha^*, R_y(\alpha)_j)]$ - if $d_f(k) = -1$ then widen the α-cut $Y\alpha_j = [L_y(\alpha_j), R_y(\alpha)_j)]$ in the right side transforming it into the new α-cut $Y'\alpha_j = [L_y(\alpha_j), y_\alpha^*]$
4	For α_j, $j = p-1, p-2, \ldots, 1$ - calculate $Rx_k^{nec}(\alpha_j)$, using (5) - if $Rx_k^{nec}(\alpha_j) < Lx_k^{nec}(\alpha_{j+1})$ then calculate y_α^* as follows: $y_\alpha^* = f(D_X(\alpha_j^*, -d_f))$ $D_X(\alpha_j^*, -d_f) = \{D_X(\alpha_{ji}^*, -d_f(i) d_f(k)$ $\mid i = 1, 2, \ldots, n\}$ - if $d_f(k) = -1$ then widen the α-cut $Y\alpha_j = [L_y(\alpha_j), R_y(\alpha)_j)]$ in the left side transforming it into the new α-cut $Y'\alpha_j = [y_\alpha^*, R_y(\alpha)_p)]$ - if $d_f(k) = 1$ then widen the α-cut $Y\alpha_j = [L_y(\alpha_j), R_y(\alpha)_j)]$ in the right side transforming it into the new α-cut $Y'\alpha_j = [L_y(\alpha_j), y_\alpha^*]$

Let us suppose that the fuzzy numbers $A_1, \ldots A_{k-1}, A_{k+1}, \ldots A_n$ represent the numerical values of the variables $x_1, \ldots, x_n$ and that each has a trapezoidal shape, and also that the fuzzy number Y represents the output variable y and has a singleton shape. We want to obtain $x_k^{nec}(\alpha)$: in other

words, we want to obtain a fuzzy number A_k (whose α-cuts are $x_k^{nec}(\alpha)$) so that it ensures that the output will be a singleton, when the other input variables are trapezoidal ones. Obviously such a fuzzy number cannot be found, because the uncertainty included in the trapezoidal numbers cannot "just disappear".

To solve this problem, we propose algorithm 3, that verifies whether the fuzzy number Y is a coherent value with the uncertainties of the input variables X_k. If it is not, then the algorithm modifies the fuzzy number Y adding uncertainty (by widening its α-cuts) in a convenient way, so that we can ensure that Y and $x_k^{nec}(\alpha)$ will be fuzzy numbers. (Steps 2, 3 and 4 may be followed in the same iteration cycle)

4.3 FAMILY OF INTERMEDIATE EXTENSIONS

By comparing equations (4) and (5), it can be observed that the main difference between the possible and necessary extensions lies in the selection of the α-cut extreme of every variable in X_k used to calculate $x_k^{pos}(\alpha)$ or $x_k^{nec}(\alpha)$. This allows us to define intermediate extensions using intermediate values of the α-cuts, by using a parameter r $(0 \leq r \leq 1)$. The possible extension will be associated with a value of $r = 0$ and the necessary solution with $r = 1$. The intermediate extensions is called $x_k^{int}(\alpha, r)$ and is calculated using a linear interpolation of the values used in (4) and (5) inside the algorithm 3. It is important to emphasize that possible extension always exists but that necessary extension does not. When necessary extension does not exist, there must be a threshold value of r, for example r_o, so that if $r < r_o$ then extension exists; r_o can then be used to measure the existence of a solution to the problem of extending the inverse function.

5 EXAMPLE

In the following example we will consider the crisp function $y = f(x_1, x_2) = \frac{x_1^2}{x_2}$ defined for strictly positive real numbers x_1,x_2,y, and their inverse function $x_1 = f_1^{-1}(y, x_2) = \sqrt{yx_2}$ If y, x_2 are represented by trapezoidal fuzzy numbers $y = T(1,1,1,1)$, $x_2 = T(0.5, 0.9, 1.1, 1.5)$ then the values of x_1 can be obtained using the family of intermediate extensions of $f_1^{-1}(y, x_2)$. The result is shown in figure 1, but y has been modified as is shown in figure 2. Note in this example how the uncertainty of x_1,x_2, changes when r changes. Note also that in this example the threshold of existence of an extension is $r_o = 0.5$, because if r takes a value larger than 0.5 the algorithm has to modify y in order to obtain a solution.

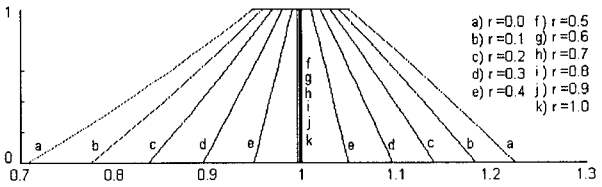


Figure 1: x_2 in example 1

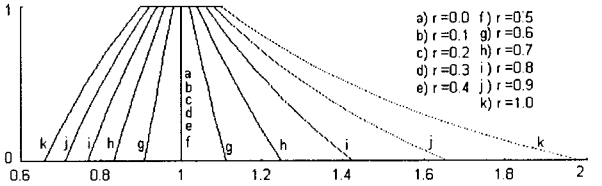


Figure 2: y in example 1

6 CONCLUSIONS AND REMARKS

In this paper we have presented some algorithms to extend crisp functions and their inverse functions to fuzzy numbers. The type of functions dealt with here are continuous functions monotonically increasing with some of their arguments, and monotonically decreasing with the others. We have also proposed some different solutions to the problem of extending crisp inverse functions: first, we proposed the possible and the necessary extensions, and then we proposed a family of intermediate extensions that varies slowly between the first two. Although some of the extensions may not exist, the algorithms proposed ensure that a solution is found, by conveniently modifying the condition of the problem.

References

[1] B. Bouchon-Meunier, O.Kosheleva, V. Kreinovich and H.T. Nguyen , Fuzzy Numbers are the only fuzzy sets that keep invertible operations invertible,*Fuzzy Sets and Systems* (1997) 155-164

[2] R. Giachetti, Evaluating Engineering functions with imprecise quantities *7th IFSA World Congress Prague* (1997) vol II 150-155

[3] R.R. Yager, On the lack of inverses in fuzzy arithmetic, *Fuzzy Sets ans Systems* (1980) 73- 82

[4] L.A. Zadeh, The concept of linguistic variable and its applications to approximate reasoning, Part I, *Inform. Sci.* 8 (1975) 199-249; Part II, *Inform. Sci.* 8 (1976) 301-357; Part III, Inform. Sci. 9 (1976) 43-80.

EMBEDDING SOME FAMILIES OF FUZZY RELATIONS IN THE CLASS OF FUZZY CONSEQUENCE OPERATORS.

Jorge Elorza[1]
Dep. de Matemática e Informática
Universidad Pública de Navarra
Campus de Arrosadía s/n
31006 Pamplona. Spain
e-mail: elorza@si.upna.es
Tel: 34948169556 Fax: 34948169521
[1] *Corresponding author*

Pedro Burillo
Dep. de Automática y Computación
Universidad Pública de Navarra
Campus de Arrosadía s/n
31006 Pamplona. Spain
e-mail: pburillo@si.upna.es
Tel: 34948169252 Fax: 34948169281

Abstract

The aim of this paper is to show several isomorphic structures between fuzzy relations and operators from the family of fuzzy sets into itself. In particular, we prove that the $*$-preorders family is able to embed itself in the class of fuzzy consequence operators by way of a lattice monomorphism.

Keywords : Fuzzy Consequence Operator, Closure Operator, Fuzzy Preorder, Fuzzy Logic, Lattice.

1 INTRODUCTION

Since L. Zadeh introduced Fuzzy Set Theory ([13], 1965), the use of the $*$-preorder concept is well-known.

Given X a nonempty universal set and given a t-norm $*$, a fuzzy relation on X (fuzzy subset of $X \times X$) is called $*$-fuzzy preorder if it verifies :
(R1) $R(x,x) = 1 \ \forall \ x \in X$ (reflexivity)
(R2) $R(x,z) \geq R(x,y) * R(y,z) \ \forall \ x,y,z \in X$ ($*$-transitivity).

In this paper Γ', Γ_r and Γ^* will represent the class of fuzzy relations on X, the family of fuzzy reflexive relations and the subfamily of fuzzy $*$-preorders, respectively.

In fuzzy logic, the fuzzy consequence operator concept is relevant, FCO abbreviately. Given L a closed lattice, a function $C : L^X \longrightarrow L^X$ is called FCO (J. Pavelka, [10]) if it verifies :
(C1) $\mu \subseteq C(\mu) \ \forall \ \mu \in L^X$ (inclusion)
(C2) $\mu_1 \subseteq \mu_2 \Longrightarrow C(\mu_1) \subseteq C(\mu_2) \ \forall \ \mu_1, \ \mu_2 \in L^X$ (monotony)
(C3) $C(C(\mu)) = C(\mu) \ \forall \ \mu \in L^X$ (idempotence).

These operators, sometimes called closure operators, are studied in a general context ([12]). During the last decade, these operators have been also studied in the context of fuzzy logic, taking the chain $L = [0,1]$ as a special case ([2],[3],[4],[6],[7],[11]). In this paper we will consider $L = [0,1]$.

We will denote by Ω' the class of operators from $[0,1]^X$ into $[0,1]^X$, Ω will represent the family of fuzzy consequence operators and for every t-norm $*$, Ω_p^* will represent the subfamily of fuzzy consequence operators which are induced by a $*$-preorder through Zadeh's max-$*$ compositional rule, this is :
$\Omega_p^* = \left\{C \in \Omega \mid \exists \, R \in \Gamma^*, \, C = C_R^*\right\}$, where $C_R^*(\mu) = \mu \circ^* R : C_R^*(\mu)(x) = \sup_{w \in X} \left\{\mu(w) * R(w,x)\right\}$.

Finally, we recall some notions and results on general lattice theory ([1],[9]) :

Given $(L, \leq, \vee, \wedge)$ a lattice and given S a subset of L, we will say that S is a sublattice of L if $(S, \leq, \vee, \wedge)$ is a lattice, this is, if it verifies :
$a \vee b \in S \ \forall \ a,b \in S$ (1) and $a \wedge b \in S \ \forall \ a,b \in S$ (2).
In this case we will write $S \leq L$. If only (1) is verified we will say that S is a $\vee$-sublattice (join-sublattice) of L and we will write $S \leq_\vee L$. If only (2) is verified we will say that S is a $\wedge$-sublattice (meet-sublattice) of L and we will write $S \leq_\wedge L$.

Let $\alpha : L \longrightarrow M$ be a function from a lattice L into a lattice M. We will say that α is isotone when $a \leq b$ implies $\alpha(a) \leq \alpha(b)$. We will say that α is a $\vee$-morphism (join-morphism) if $\alpha(a \vee b) = \alpha(a) \vee \alpha(b)$ for all $a, b \in L$; a $\wedge$-morphism (meet-morphism) if $\alpha(a \wedge b) = \alpha(a) \wedge \alpha(b)$ for all $a, b \in L$; and a morphism when both conditions hold. As always, a morphism is called a monomorphism if it is one-one; an epimorphism if onto and an isomorphism if a bijection. If there exists an isomorphism between L and M, we will write $L \cong M$; if there exists a join-isomorphism, $L \cong_{\vee} M$ and if a meet-isomorphism, $L \cong_{\wedge} M$.

The following result holds : Any join-morphism between lattices is isotone; also any meet-morphism is isotone; and finally, any isotone bijection with isotone inverse is a lattice isomorphism.

2 LATTICE SUBSTRUCTURES

It is well-known (J.A. Goguen, [8]) that if L is a distributive complete lattice then the functional space L^X has the same structure.

As $([0,1], \leq, \vee, \wedge)$ is a distributive complete lattice (in fact, it is a chain), so is the family $[0,1]^X$ of fuzzy subsets of X with the following natural operations :

$$\mu_1 \subseteq \mu_2 \iff \mu_1(x) \leq \mu_2(x) \ \forall\, x \in X$$
$$(\mu_1 \vee \mu_2)(x) = \max\left(\mu_1(x), \mu_2(x)\right)$$
$$(\mu_1 \wedge \mu_2)(x) = \min\left(\mu_1(x), \mu_2(x)\right).$$

The family $\Gamma' = [0,1]^{X \times X}$ as family of fuzzy subsets of $X \times X$ is also a distributive complete lattice with the previous operations between fuzzy relations.

Analogously the family $\Omega' = \left([0,1]^X\right)^{[0,1]^X}$ is another distributive complete lattice with the operations :

$$C_1 \subseteq C_2 \iff C_1(\mu) \leq C_2(\mu) \ \forall\, \mu \in [0,1]^X$$
$$\left(C_1 \vee C_2\right)(\mu)(x) = \max\left(C_1(\mu)(x), C_2(\mu)(x)\right)$$
$$\left(C_1 \wedge C_2\right)(\mu)(x) = \min\left(C_1(\mu)(x), C_2(\mu)(x)\right).$$

On the other hand, it is clear that $\Gamma_r \leq \Gamma'$ and we proved outside this work that $\Gamma^* \leq_{\wedge} \Gamma'$ and $\Omega_p^* \leq_{\wedge} \Omega'$ but $\Gamma^* \not\leq_{\vee} \Gamma'$ and $\Omega_p^* \not\leq_{\vee} \Omega'$, for every t-norm $*$. Thus:

$$\Gamma^* \leq_{\wedge} \Gamma_r \leq \Gamma' \quad \text{and} \quad \Omega_p^* \leq_{\wedge} \Omega \leq \Omega'.$$

3 ISOMORPHIC STRUCTURES

For every t-norm $*$, we consider the function $\theta'^* : \Gamma' \longrightarrow \Omega'$ given by $\theta'^*(R) = C_R^*$ for any fuzzy relation R, where $C_R^*(\mu)(x) = \sup_{w \in X} \left\{\mu(w) * R(w,x)\right\}$. θ^* will represent the restriction of θ'^* to the family Γ^*. We also consider the function $\tilde{\theta}' : \Omega' \longrightarrow \Gamma'$ given by $\tilde{\theta}'(C) = R_C$ for any operator C, where $R_C(x,y) = C(\varphi_x)(y)$ and φ_x represents the crisp membership of the singleton $\{x\}$. We can note that the function $\tilde{\theta}'$ does not depend on the t-norm $*$.

The following result holds for any fuzzy relation R. This result is proved by J.L. Castro and E. Trillas ([4]) if we suppose that R is a fuzzy $*$-preorder.

Theorem 1.- *For any fuzzy relation R on X and for any t-norm $*$, $R_{C_R^*}$ is exactly the relation R.*

Proof.- $R_{C_R^*}(x,y) = C_R^*(\varphi_x)(y)$. But

$$C_R^*(\varphi_x)(y) = \sup_{w \in X} \left\{\varphi_x(w) * R(w,y)\right\}$$

If $w \neq x$, $\varphi_x(w) = 0$. Hence this supremum is taked in $w = x$ and it is equal to

$$\varphi_x(x) * R(x,y) = 1 * R(x,y) = R(x,y) \qquad \square$$

As a consequence :

Corollary 1.- *$\tilde{\theta}' \circ \theta'^*$ results the identity mapping on Γ'.*

Corollary 2.- *The function $\theta'^* : \Gamma' \longrightarrow \Omega'$ is one to one.*

Proof.- If $\theta'^*(R_1) = \theta'^*(R_2)$, $\tilde{\theta}' \circ \theta'^*(R_1) = \tilde{\theta}' \circ \theta'^*(R_2)$ and from corollary 1, $R_1 = R_2$. $\square$

We will denote by $\Omega_p'^*$ the set $\Im m(\theta'^*)$. It is clear that $\Omega_p'^* \subsetneq \Omega'$ and thus θ'^* is not onto. In fact :

Theorem 2.- $\Omega_p'^* = \left\{C \in \Omega' \mid \theta'^* \circ \tilde{\theta}'(C) = C\right\}$.

Proof.- If $C \in \Im m(\theta'^*)$, there exists a unique $R \in \Gamma'$ such that $\theta'^*(R) = C$. From corollary 1, $\tilde{\theta}' \circ \theta'^*(R) = R$ and $\theta'^* \circ \tilde{\theta}'(C) = (\theta'^* \circ \tilde{\theta}') \circ \theta'^*(R) = \theta'^* \circ (\tilde{\theta}' \circ \theta'^*)(R) = \theta'^*(R) = C$. Conversely, if $\theta'^* \circ \tilde{\theta}'(C) = C$, C is the image of the fuzzy relation $\tilde{\theta}'(C)$ by the mapping θ'^* and thus $C \in \Im m(\theta'^*)$. $\square$

Theorem 3.- *The function $\theta'^* : \Gamma' \longrightarrow \Omega_p'^*$ is a lattice isomorphism. Thus : $\Gamma' \cong \Omega_p'^*$ and $\Omega_p'^* \leq \Omega'$.*

Proof.- θ'^* is obviously an isotone bijection. But $\theta'^{*-1} : \Omega_p'^* \longrightarrow \Gamma'$ is exactly the restriction of $\tilde{\theta}'$ to $\Omega_p'^*$ which is also isotone. Therefore θ'^* is a lattice isomorphism. $\square$

Accordingly, we recall that the function θ'^* is a lattice morphism : $\theta'^*(R_1 \vee R_2) = \theta'^*(R_1) \vee \theta'^*(R_2)$ and $\theta'^*(R_1 \wedge R_2) = \theta'^*(R_1) \wedge \theta'^*(R_2)$ for all fuzzy relations R_1, R_2 .

In particular, we have for all $*$-preorders R_1, R_2 : $\theta'^*(R_1 \vee R_2) = \theta^*(R_1) \vee \theta^*(R_2)$ and $\theta^*(R_1 \wedge R_2) = \theta^*(R_1) \wedge \theta^*(R_2)$, where θ^* is a bijection between Γ^* and Ω_p^* with inverse isotone.

However, since Γ^* is not a sublattice of Γ' we cannot say that there exists an isomorphism between Γ^* and Ω_p^* with the usual lattice operations. But it is possible to modify the previous operations to obtain an isomorphism between both families. We will analyze this question at the end of this section.

At all events, we have the following representation theorems :

Theorem 4.- *The function $\theta^* : \Gamma^* \longrightarrow \Omega_p^*$ is a meet-isomorphism. Thus : $\Gamma^* \cong_\wedge \Omega_p^* \leq_\wedge \Omega \leq_\wedge \Omega'$.*

Theorem 5.- *The restricted function $\theta'^* \mid \Gamma_r : \Gamma_r \longrightarrow \Omega'$ is a lattice monomorphism. Thus :*
$\Gamma_r \cong \Omega_{pi}'^*$, where $\Omega_{pi}'^* = \Im m_{\theta'^*}(\Gamma_r)$, this is :
$\Omega_{pi}'^* = \left\{C \in \Omega_p'^* \mid C \text{ verifies the inclusion axiom}\right\}$.

We will show a last isomorphism with the usual operations, extending minimally the family Γ^* for this purpose :

Let $\overline{\Gamma}^*$ be the sublattice generated by Γ^*, this is, the smallest sublattice of Γ' containing Γ^*. It is clear that $\overline{\Gamma}^* = \left\{R_1 \vee R_2 \vee \cdots \vee R_n \mid n \in \mathbb{N}, R_i \in \Gamma^*\right\}$.

Now, we consider $\overline{\theta}^*$ as the restricted function θ'^* to $\overline{\Gamma}^*$. Analogously, $\overline{\theta}^*$ is a lattice monomorphism. Moreover, if we denote by $\overline{\Omega}_p^*$ the set $\Im m(\overline{\theta}^*)$ then $\overline{\Omega}_p^* = \left\{C_1 \vee C_2 \vee \cdots \vee C_n \mid n \in \mathbb{N}, C_i \in \Omega_p^*\right\}$ and $\overline{\Omega}_p^*$ result the sublattice of Ω' generated by Ω_p^*. Hence :

Theorem 6.- *The function $\overline{\theta}^* : \overline{\Gamma}^* \longrightarrow \overline{\Omega}_p^*$ is a lattice isomorphism. Thus : $\overline{\Gamma}^* \cong \overline{\Omega}_p^* \leq \Omega'$.*

The previous result is a representation theorem of all fuzzy relations which are finite union of fuzzy $*$-preorders by way of the family of operators which are finite union of fuzzy consequence operators.

From now, we will consider $*$ as a **continuous** t-norm with the aim of obtaining an isomorphism between Γ^* and Ω_p^*.

Since Γ^* is not a join-sublattice of Γ' we are obliged to vary the $\vee$ operation for this purpose to be possible.

Given a continuous t-norm $*$, it is proved ([3]) that the family Γ^* of all fuzzy $*$-preorders is a complete lattice with the usual $\subseteq$, $\wedge$ operations and the following $\sqcup^*$:

$$R_1 \sqcup^* R_2 = \bigcap \left\{R \in \Gamma^* \mid R \supseteq R_1 \vee R_2\right\}.$$

Notice that the following properties hold, hence $\sqcup^*$ can work as a union operation :
($\sqcup^*$ 1) $R_1 \sqcup^* R_2 \supseteq R_1 \vee R_2$.
($\sqcup^*$ 2) $R_1 \sqcup^* R_2 = R_1 \vee R_2 \iff R_1 \vee R_2 \in \Gamma^*$.
($\sqcup^*$ 3) $R_1 \sqcup^* R_2$ is the smallest $*$-preorder containing R_1 and R_2.
($\sqcup^*$ 4) $R_1 \sqcup^* R_2 = R_1 \iff R_1 \supseteq R_2$.

The next example proves that $\left(\Gamma^*, \sqcup^*, \wedge, \subseteq\right)$ is a **non**-distributive lattice.

Let X be an universe with card$X \geq 3$, we consider the following crisp relations on X :

$$R_1 = \left\{(x,y),(y,x)\right\} \cup \left\{(t,t)\right\}_{t \in X}$$
$$R_2 = \left\{(y,z),(z,y)\right\} \cup \left\{(t,t)\right\}_{t \in X}$$
$$R_3 = \left\{(x,z),(z,x)\right\} \cup \left\{(t,t)\right\}_{t \in X}$$

We have : R_1, R_2, $R_3 \in \Gamma^*$, $R_1 \vee R_3$, $R_2 \vee R_3 \notin \Gamma^*$. So $R_1 \wedge R_2 = \left\{(t,t)\right\}_{t \in X} \in \Gamma^*$. On the other hand $R_1 \sqcup^* R_3 = R_2 \sqcup^* R_3 = R_1 \vee R_2 \vee R_3$. Thus :

$$\left(R_1 \sqcup^* R_3\right) \wedge \left(R_2 \sqcup^* R_3\right) = R_1 \vee R_2 \vee R_3 \neq$$

$$\neq R_3 = \left(R_1 \sqcup^* R_2\right) \wedge R_3$$

Now, we define analogously the operation $\sqcup^*$ in Ω_p^*. Given C_1, $C_2 \in \Omega_p^*$:

$$C_1 \sqcup^* C_2 = \bigcap \left\{C \in \Omega_p^* \mid C \supseteq C_1 \vee C_2\right\}$$

This operation is well defined : let R be the trivial relation $R(x,x) = 1$ for all $x \in X$ then for any t-norm $*$, $\theta^*(R)$ is the operator C_{sup} given by

$$C_{sup}(\mu)(t) = \sup_{w \in X} \left\{\mu(w)\right\}.$$

Clearly, C_{sup} is the greatest element of the set Ω_p^*. Hence, given C_1, $C_2 \in \Omega_p^*$, $C_{sup} \supseteq C_1 \vee C_2$ and thus $\left\{C \in \Omega_p^* \mid C \supseteq C_1 \vee C_2\right\}$ is a nonempty set. As Ω_p^* is a complete $\wedge$-sublattice, the operation $\sqcup^*$ is well defined.

The analogous properties ($\sqcup^*$ 1),($\sqcup^*$ 2),($\sqcup^*$ 3),($\sqcup^*$ 4) for $*$-preorders also hold for operators. Hence $\sqcup^*$ also can work as a union operation in Ω_p^*.

It is easy to prove that $\left(\Omega_p^*, \sqcup^*, \wedge, \subseteq\right)$ is also a complete **non**-distributive lattice.

Finally, we can show the representation theorem which identifies both structures.

Theorem 7.- *The function θ^* is a lattice isomorphism between* $\left(\Gamma^*, \sqcup^*, \wedge, \subseteq\right)$ *and* $\left(\Omega_p^*, \sqcup^*, \wedge, \subseteq\right)$.

Proof.- The bijection θ^* from Γ^* into Ω_p^* is isotone and so its inverse $\theta^{*^{-1}} = \tilde{\theta} \mid \Omega_p^*$. □

4 CONCLUSIONS

It is clear that the functions θ'^* change essentially on t-norm $*$. Consequently, the set $\Omega_p^{'*}$ of operators $*$-induced by relations is different from $\Omega_p^{'\perp}$ if $* \neq \perp$.

Nevertheless, by theorem 3, $\Omega_p^{'*} \cong \Omega_p^{'\perp} \cong \Gamma'$ for all t-norms $*$, $\perp$. Analogously, by theorem 5, $\Omega_{pi}'^* \cong \Omega_{pi}'^{\perp} \cong \Gamma_r$.

It is not true for two sets Ω^* and $\Omega^\perp$ of preorders induced fuzzy consequence operators, this is, Ω^* and $\Omega^\perp$ are not isomorphic if $* \neq \perp$.

Moreover, we have shown the minimal isomorphism $\overline{\Gamma}^* \cong \overline{\Omega}_p^*$ with the usual $\vee$, $\wedge$ operations, where $\overline{\Gamma}^*$ is the sublattice generated by the family of $*$-fuzzy preorders.

In the context of fuzzy preorders and fuzzy consequence operators we have only a meet-isomorphism : $\Gamma^* \cong_\wedge \Omega_p^*$.

We had to change the usual union to obtain an isomorphism between Γ^* and Ω_p^*. In fact, we show that θ^* is a lattice isomorphism between the non-distributive lattices $\left(\Gamma^*, \sqcup^*, \wedge, \subseteq\right)$ and $\left(\Omega_p^*, \sqcup^*, \wedge, \subseteq\right)$.

We recall that the class Ω of fuzzy consequence operators is not a sublattice in Ω'. However, Ω form a lattice within Ω' with the usual intersection and the following union ([2],[12]) :

$$C_1 \sqcup' C_2 = \bigcap \{C \in \Omega \mid C \supseteq C_1 \vee C_2\} \ \forall\, C_1, C_2 \in \Omega.$$

We proved recently ([5]) that the restriction of $\sqcup'$ to the family Ω_p^* is precisely $\sqcup^*$. Therefore we have embedded the family of all $*$-preorders Γ^* in the class of fuzzy consequence operators Ω by a lattice monomorphism : $\Gamma^* \cong \Omega_p^* \leq \Omega$.

Finally, since the lattice Ω and the lattice of all closure systems on X (a family of fuzzy subsets of X is called closure system when it is closed under arbitrary intersections and it contains X) are dually isomorphic ([2]), the families Γ^* and Ω_p^* are called to embedding in the structure of closure systems under the dual isomorphism concept.

The purpose of this paper is to complete the relation between fuzzy preorders and FCOs in order to use the FCO concept in approximate reasoning. This question will require further research.

Acknowledgement

We thank the referees for their valuable suggestions that have improved the final release of this paper.

References

[1] Birkhoff, G. (1967) *Lattice Theory.* 3rd ed. American Mathematical Society Colloquium Publications, Vol 25.

[2] Castro, J.L. (1994) *Fuzzy logics as families of bivaluated logics.* Fuzzy sets and systems 64, pg. 321-332.

[3] Castro, J.L. - Delgado, M. - Trillas, E. (1994) *Inducing Implication Relations.* International Journal of Approximate Reasoning 10, pg. 235-250

[4] Castro, J.L. - Trillas, E. (1991) *Tarski's fuzzy consequences*, submit to IJAR (Proc. of the Internat. Fuzzy Engineering Symp'91. Vol.1, pg. 70-81.)

[5] Elorza, J. (1999) *Generación, Estructuras y Representación de los Operadores de Consecuencias de la Lógica Borrosa.* Tesis Doctoral. Universidad Pública de Navarra.

[6] Esteva, F. - García, P. - Godo, L. - Rodríguez, R.O. (1996) *Fuzzy approximation relations, Modal structures and possibilistic logic.* Submit to Mathware & Soft Computing.

[7] Gerla, G. (1994) *An Extension Principle for Fuzzy Logics.* Mathematical Logic Quartely, 40, pg. 357-380.

[8] Goguen, J.A. (1969) *The logic of inexact concepts.* Synthese 19, pg. 325-373.

[9] Grätzer, G. (1978) *General Lattice Theory.* Ed. Birkhäuser.

[10] Pavelka, J. (1979) *On Fuzzy Logic I.* Zeitschr. f. Math. Logik und Grundlagen d. Math. Bd.25, pg. 45-52.

[11] Trillas, E. - Alsina, C. - Terricabras, J.M. (1995) *Introducción a la lógica borrosa.* Ed.Ariel.

[12] Ward, M. (1940) *The closure operators of a lattice.* Annals of Mathematics. Vol.43. N.2, pg. 191-196.

[13] Zadeh, L.A. (1965) *Fuzzy sets.* Information and control, 8, pg. 338-353.

FUZZY ARITHMETIC AND FUZZY RANKING: A NONSTANDARD APPROACH

Luis Marín
Departamento de Matemáticas
Universidad de Cádiz
luis.marin@uca.es

Francisco Mesa
Departamento de Ingeniería Eléctrica
Universidad de Cádiz
francisco.mesa@uca.es

Summary

In this paper, we analyze relations between nonstandard analysis and fuzzy sets theory. We propose an interpretation of fuzzy sets using a nonstandard approach and we propose a fuzzy ranking method based on this interpretation.

Palabras Clave: Nonstandard Analysis, Ranking of fuzzy numbers..

1 Introduction

Although it seems that Cantorian set theory has prevailed in the whole field of mathematics, there are some alternative theories where the invention of new set-like concepts is inevitable. Drossos summarizes a partial list of such instances. Nonstandard Analysis (Robinson 1961) and Fuzzy Sets (Zadeh 1965) are included in this list.

We may divide these theories into two categories: The extensional and the intentional one. Let A be a set, then any function

$$p : I \to A \tag{1}$$

represents an extensional aspects of A (e.g. a list of elements, a path, etc.). If the index set is $I = \{1, 2, \ldots, n\}$, every application p can be associated with an element of A^n. Nonstandard analysis is another example of extension.

Coming now to intentional aspects of a set A, we consider instead of incoming functions, outgoing ones. That is functions of the type

$$p : A \to I. \tag{2}$$

If the index set is $I = \{1, 2, \ldots, n\}$, we divide the elements of A according to a given property. The definition of a fuzzy subset of A is another example of intention.

2 Nonstandar Analysis Development

The classic way of constructing hyper-real numbers is to consider the natural numbers set N, an ultrafilter U in N (usually an ultrafilter containing Frechets filter) and the real numbers set R. Considering the set of applications from N to R and the equivalence classes given by the equivalence relation

$$f \sim g \quad \text{iff} \quad \{n \in N : f(n) = g(n)\} \in U, \tag{3}$$

the set of the hyper-real numbers is obtained. This set constitutes an ordered field, and it is non-archimedian since it contains numbers which are greater than all numbers of R , which constitutes a subfield of it.

In this paper we consider the construction of the hyper-reals as follows: we use the set $I = [0, 1]$ instead of N and an ultrafilter containing the filter formed by all sets which the Lebesgue's measure is equal to 1. Considering an equivalence relation as in (3)

$$f \sim g \quad \text{iff} \quad \{x \in [0, 1] : f(x) = g(x)\} \in U, \tag{4}$$

we have an ordered field which we call HR.

The addition and multiplication of hyper-real numbers are defined in a natural way as follows:

$$(f + g)(x) = f(x) + g(x) \tag{5}$$

$$(fg)(x) = f(x)g(x) \tag{6}$$

Where the functions defined as

$$f(x) = 0 \quad \forall x \in I \qquad f(x) = 1 \quad \forall x \in I \tag{7}$$

are, respectively, the neutral elements for addition and multiplication. The associativity and commutativity

can be trivially probed for both operators. The inverse element of f under addition is $-f$ and, if f is a non-zero element, the function $1/f$ defined as

$$(1/f)(x) = \begin{cases} 1/f(x) & \text{if} \quad f(x) \neq 0 \\ 1 & \text{if} \quad f(x) = 0 \end{cases} \tag{8}$$

is the inverse element of f under multiplication. This affirmation can be shown calculating

$$\Big((f)(1/f) \Big)(x) = 1 \quad \forall x \in P \tag{9}$$

where

$$P = \{x : f(x) \neq 0\} \in U \tag{10}$$

If the set P shown above is not an element of U then f is in the equivalence class of zero. Therefore f is in the equivalence class of 1.

The usual definition of an order relation in HR is

$$f < g \text{ iff } \{x \in I : f(x) < g(x)\} \in U$$

With this relation, HR is a totally ordered field.

We have two dual constructions:

- a function from $I = [0,1]$ to R is a member of the hyper-real numbers field (with an appropriate equivalence relation).
- an application from R to $I = [0,1]$ is a fuzzy set.

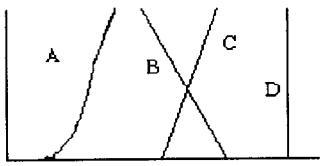


Figure 1: Hyper-real numbers

Fig. 1 shows four different elements of HR, named A, B, C and D. In all the graphics we use, we represent the set I vertically and the set R horizontally. The hyper-real number D, corresponding to the function

$$f_D(x) = d \quad \forall x \in I,$$

can be interpreted as the representation of the real number d.

From the definition of '$<$', we obtain the following relations: $A < B$, $A < C$, $A < D$, $B < D$ and $C < D$. The comparison between B and C is not trivial, depending on the considered ultrafilter U. With one ultrafilters we can obtain $B < C$ and with another $C < B$.

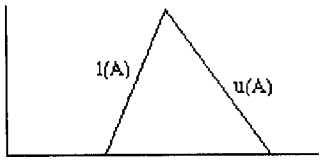


Figure 2: Triangular fuzzy number.

Like we can not find an analytic definition of U, the total ordering defined above has a limited use.

We propose to define a fuzzy relation R to compare hyper-real numbers as follows:

$$R(B,C) = \mu\Big(\{x \in [0,1] : B(x) < C(x)\} \Big) \tag{11}$$

for any measure μ defined in $[0,1]$. This fuzzy relation can be interpreted as the proportion of ultrafilters in a family verifying $B < C$.

3 A Nonstandard interpretation of fuzzy numbers

In the graphic representation shown above, we can see how hyper-real numbers and fuzzy numbers have a dual nature. A fuzzy number we call in this paper as follows

Definition A fuzzy number A is defined as a fuzzy set on the space of real numbers R, whose membership function μ_A satisfies the following conditions:

1. μ_A is a mapping from R to the closed interval $[0,1]$,
2. there exists a unique real number m such that
   (a) $\mu_A(m) = 1$
   (b) μ_A is nondecreasing on $(-\infty, m]$,
   (c) μ_A is nonincreasing on $[m, +\infty)$,

In the above definition we should notice that no continuity is assumed to the membership function. We denote the set of all fuzzy numbers by FR.

Figure 2 shows a classic representation of a triangular fuzzy number. We can see that fuzzy number A is limited by two functions: the lower bound $l(A)$ and the upper bound $u(A)$. Every α-level cut can be expressed as $(l(A)(\alpha), u(A)(\alpha))$. But the functions $l(A)$ and $u(A)$ can also be interpreted as hyper-real numbers, verifying $l(A) < u(A)$.

Analyzing this fact, we interpret a fuzzy number as an interval in the set of hyper-real numbers: The set of

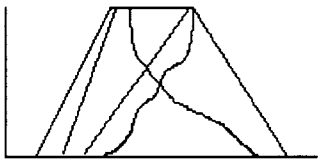


Figure 3: Trapezoid fuzzy number

all hyper-real numbers h verifying $l(A) < h < u(A)$. Figure 3 shows this idea using a trapezoid fuzzy number.

Although we can consider every fuzzy set as an interval of hyper-real numbers, the opposite implication is not true. If we consider two hyper-real numbers f and g verifying $f(0) = g(0)$, $f(1) = g(1)$ and $f(x) < g(x) \forall x \in (0,1)$, obviously $f < g$ and they are the bounds of an interval of hyper-real numbers, but they cannot be interpreted as the lower bound and the upper bound of a fuzzy set.

The addition and multiplication of fuzzy numbers is deduced from interval arithmetic. Assume that A and B are two fuzzy numbers, which bounds are $(l(A), u(A))$ and $(l(B), u(B))$ respectively. The addition of A and B is the fuzzy number C which bounds are $(l(A) + l(B), u(A) + u(B))$.

The lower bound of the resulting fuzzy set C is the sum of the lower bounds of the fuzzy numbers A and B, and the upper bound of the resulting fuzzy set C is the sum of the upper bounds of the fuzzy numbers A and B. Using α-level sets we obtain the same result. The multiplication of A and B is the fuzzy number which bounds are $(l(A)l(B), u(A)u(B))$.

4 Ranking fuzzy numbers

Using this interpretation of fuzzy sets, we will define a fuzzy relation L with the purpose of ranking fuzzy numbers. At first time, it is necessary the definition of a measure of intervals $\mu(A)$. In this paper, we consider the measure of an interval $A = (l(A), u(A))$ given by the area below the two bounded curves. If $A_i \subset A$, a relative measure may be defined as follows:

$$M_A(Ai) = \frac{\mu(Ai)}{\mu(A)} \tag{12}$$

The fuzzy relation we are building must verify the conditions:

$$L(A,B) = 1 \text{ if } \max(A) < \min(B) \tag{13}$$

$$L(A,B) = 0 \text{ if } \min(A) < \max(B) \tag{14}$$

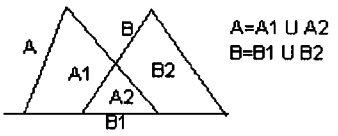


Figure 4: Comparison of fuzzy numbers

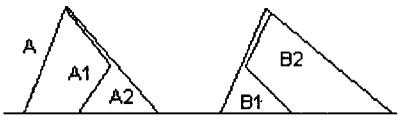


Figure 5: Making subintervals

$$L(A,A) = 1/2 \tag{15}$$

where max and min are interpreted as the maximum and the minimum of all hyper-real numbers in the interval. The last condition is introduced to ensure reciprocity and weak reflexivity.

We considered that several properties related to partitions must be verified by this relation. If it is considered partitions of the sets

$$A = A_1 \cup A_2 \cup \ldots \cup A_n \text{ with } A_i \cap A_j = \emptyset \text{ if } i \neq j$$

$$B = B_1 \cup B_2 \cup \ldots \cup B_m \text{ with } B_i \cap B_j = \emptyset \text{ if } i \neq j$$

The fuzzy relation we are building must verify the next properties:

$$L(A,B) = \sum_{i=1}^{n} M_A(A_i) L(A_i, B) \tag{16}$$

and

$$L(A,B) = \sum_{j=1}^{m} M_B(B_j) L(A, B_j) \tag{17}$$

therefore

$$L(A,B) = \sum_{i=1}^{n} \sum_{j=1}^{m} M_A(A_i) M_B(B_j) L(A_i, B_j) \tag{18}$$

Whith an appropriated selection of the partitions, $L(A_i, B_j)$ can be calculated using the conditions and properties in a great number of situations. The rest of the cases will be discussed bellow. These cases occur when the intersection area and the areas of both sets are the same.

For instance, figure 4 shows two fuzzy numbers with $supp(A) \cap supp(B) \neq \emptyset$, Number $A = (l(A), u(A))$ can

be divided in two subintervals A_1 and A_2, and number B can be divided in B_1 and B_2, as shown in figure 5.

We define a new hyperreal number $m(A)$ verifying $m(A)(x) = \min[u(A)(x), l(B)(x)]$.

So $A_1 = (l(A), m(A))$ and $A_2 = (m(A), u(A))$.

By analogy, $m(B)(x) = \max[u(A)(x), l(B)(x)]$, and so $B_1 = (l(B), m(B))$ and $B_2 = (m(B), u(B))$

Like $\max(A_1) < min(B_1)$, and $\max(A_1) < min(B_1)$ then $L(A_1, B_1) = L(A_2, B_2) = 1$. By the way, $\max(A_2) < min(B_2)$, then $L(A_2, B_2) = 1$.

$L(A_2, B_1)$ is the only case unable to calculate applying the conditions (13-18). It happens because the intersection area and the areas of both sets are the same. It can be defined as $L(A_2, B_1) = 1/2$, but we prefer a more elaborated construction. We call (a, h) to the intersection point between $u(a)$ and $l(b)$. If f and g are hyper-real numbers belong to A_2 and B_1 respectively, then we obtain $f(x) < g(x) \forall x \in (h, 1]$ however, if $x \in [0, h)$, it may be happen that $f(x) < g(x)$ or $g(x) < f(x)$.

If we want to calculate $L(A, B)$ and we are in a case unable to calculate applying the conditions (13-18), we define a real function $r_{A,B}[0, 1] \longrightarrow [0, 1]$ as follows:

$$r_{A,B}(x) = \begin{cases} 1 & \text{if } f(x) < g(x) \begin{cases} \forall f \in A \\ \forall g \in B \end{cases} \\ 0 & \text{if } g(x) < f(x) \begin{cases} \forall f \in A \\ \forall g \in B \end{cases} \\ 1/2 & \text{otherwise} \end{cases} \quad (19)$$

The average of the height of $r_{A,B}$, which is equal to its area, is the value of $L(A, B)$ we define. Thus

$$L(A, B) = \int_0^1 r_{A,B}(x) dx \quad (20)$$

wich is a generalization of (11).

In the example of figures 4 and 5, the function r_{A_2,B_1} is equal to

$$r_{A_2,B_1}(x) = \begin{cases} 1 & \text{if } x \in (h, 1] \\ 1/2 & \text{if } x \in [0, h) \end{cases}$$

and the value of $L(A_2, B_1)$ is $1 - h/2$. So

$$\begin{aligned} L(A, B) &= L(A_1, B_1) M_A(A_1) M_B(B_1) + \\ &+ L(A_1, B_2) M_A(A_1) M_B(B_2) + \\ &+ L(A_2, B_1) M_A(A_2) M_B(B_1) + \\ &+ L(A_2, B_2) M_A(A_2) M_B(B_2) \end{aligned}$$

and

$$\begin{aligned} L(A, B) &= M_A(A_1) M_B(B_1) + M_A(A_1) M_B(B_2) + \\ &+ (1 - h/2) M_A(A_2) M_B(B_1) + M_A(A_2) M_B(B_2) \end{aligned}$$

Therefore

$$L(A, B) = 1 - h/2 M_A(A_2) M_B(B_1)$$

It must be noticed that $L(B_1, A_2) = h/2$. Consequently, $L(A, B) + L(B, A) = 1$ which is the reciprocity property.

5 Conclusions

The interpretation of fuzzy numbers as intervals in HR can inspire new definitions of the basic concepts. In this paper, we propose a fuzzy ranking method based on this interpretation which can be used to solve optimization problems.

Acknowledgements

The authors would like to thank the referees for their valuable comments.

References

[1] Drossos, C. Foundations of Fuzzy Sets. A nonstandard approach. *Fuzzy Sets and Systems* 1996.

[2] Furukawa, N. Parametrics Orders on Fuzzy Numbers and their roles in Fuzzy Optimization Problems. *Optimization* 1997.

[3] González, A. Números difusos, Aritmética y comparación. *Algunos Aspectos del Tratamiento de la Información en Inteligencia Artificial. Universidad de Granada* 1991.

[4] Jiménez, M. Ordenación de números borrosos mediante la comparación de sus intervalos esperados. *Actas del Cuarto Congreso de la Asociación Esp de Tecnologías y Lógica Fuzzy* 1994.

[5] Kandel, J. Fuzzy Mathematical Techniques with Applications. *Addison-Wesley* 1986.

[6] Lai, Y. J. y Hwang, C. L. Fuzzy Mathematical Programming, Methods and Applications. *Springer-Verlag* 1992.

[7] Pedrycz, W. Two Nonparametric Models for Fusing Heterogeneous Fuzzy Data. *IEEE Trans. On Fuzzy Systems* Vol. 6, No. 3. 1998.

[8] Robinson, A. Non-Standard Analysis. *Studies in Logic and the Foundations of Mathematics, North-Holland* 1966.

[9] Zadeh, L. A. Similarity Relations and Fuzzy Orderings. *Information Sciences* 3. 1971.

EXCHANGEABILITY OF EXPECTATION, DIFFERENTIAL AND SUPPORT FUNCTION

L.J. Rodríguez-Muñiz
Dto. de Estadística e I.O. y D.M.
Universidad de Oviedo
luisj@pinon.ccu.uniovi.es

M. López-Díaz
Dto. de Estadística e I.O. y D.M.
Universidad de Oviedo
mld@pinon.ccu.uniovi.es

M.A. Gil
Dto. de Estadística e I.O. y D.M.
Universidad de Oviedo
angeles@pinon.ccu.uniovi.es

Summary

In this work we present sufficient conditions to guarantee the exchangeability among the expected value, the support function and the differential of fuzzy-valued mappings depending on a real-valued parameter.

Keywords: Differentiability, Fuzzy Expected Value, Fuzzy Random Variable, Support Function.

1 INTRODUCTION

There exists a vast literature about the differentiability of fuzzy-valued mappings, see for instance Dubois and Prade [2], Puri and Ralescu [5], Goetschel and Voxman [3] or Kaleva [4]. One of the main problems in working with differentials of fuzzy-valued mappings is that the usual class of fuzzy sets of R^p does not configure a vectorial space with respect to the sum and the product by a scalar induced by Zadeh's Extension Principle [9]. Due to this problem, it becomes sometimes useful to embed that class of fuzzy sets into a normed space, in order to define a classical Fréchet-type differential. This tool has been used in Puri and Ralescu [6] and, recently, in Rojas-Medar *et al.* [8] to embed special classes of fuzzy sets into the space of continuous functions from $[0,1] \times S^{p-1}$ to R.

In order to develop an integration theory for fuzzy-valued random variables, an important question arises in the non-separability of the usual class of fuzzy sets. This problem can be avoided by using the previously referred embedding, because fuzzy-valued mappings become functions taking on values in a compact Banach space.

In some problems in Statistics it is necessary to exchange the expectation of a random variable and the differential with respect to some real-valued parameter. We can combine the above techniques in order to give an answer to this problem of exchangeability in the previous context.

2 PRELIMINARIES

Let $\mathcal{K}_c(\mathrm{R}^p)$ be the class of non-empty compact convex subsets of R^p. The well-known Hausdorff metric on this class is defined by

$$d_H(A,B) = \max\{\sup_{a\in A}\inf_{b\in B}\|a-b\|, \sup_{b\in B}\inf_{a\in A}\|a-b\|\}$$

for all $A, B \in \mathcal{K}_c(\mathrm{R}^p)$.

We will denote by $\mathcal{F}_{cc}(\mathrm{R}^p)$ the class of the fuzzy sets $V : \mathrm{R}^p \to [0,1]$ with $V_\alpha \in \mathcal{K}_c(\mathrm{R}^p)$, for all $\alpha \in [0,1]$, and the mapping $\alpha \to V_\alpha$ being continuous with respect to the Hausdorff metric (where the 0-level is intended to be the closed convex hull of the support of V).

On $\mathcal{F}_{cc}(\mathrm{R}^p)$ we can define the metric (Puri and Ralescu [7]) given by

$$d_\infty(A,B) = \sup_{\alpha\in[0,1]} d_H(A_\alpha, B_\alpha)$$

for all $A, B \in \mathcal{F}_{cc}(\mathrm{R}^p)$.

We will denote by $\mathcal{F}_{cl}(\mathrm{R}^p)$ the subclass of fuzzy sets $V \in \mathcal{F}_{cc}(\mathrm{R}^p)$ such that the mapping $\alpha \to V_\alpha$ is Lipschitz.

Puri and Ralescu [6] proved that the metric space $(\mathcal{F}_{cl}(\mathrm{R}^p), d_\infty)$ can be embedded into the space $(C([0,1]\times S^{p-1}), \|\cdot\|_\infty)$ by means of the support function of the fuzzy sets, that is, the mapping

$$j : \mathcal{F}_{cl}(\mathrm{R}^p) \to C([0,1]\times S^{p-1})$$

given by $j(A) = s_A$, for every $A \in \mathcal{F}_{cl}(\mathrm{R}^p)$ (where s_A is the support function of A) is an isometrical embedding. Recently, Rojas-Medar *et al.* [8] have extended this embedding to the class $\mathcal{F}_{cc}(\mathrm{R}^p)$.

Let $(\Omega, \mathcal{A}, P)$ a probability space. A set-valued mapping $X : \Omega \to \mathcal{K}_c(\mathrm{R}^p)$ is said to be a *random set* if it is Borel-measurable when we consider the Hausdorff metric on $\mathcal{K}_c(\mathrm{R}^p)$. A random set is said to be *integrably bounded* if there exists a mapping $h : \Omega \to \mathrm{R}$ with $h \in L^1(\Omega, \mathcal{A}, P)$, such that

$$\sup_{x \in X(\omega)} \|x\| \leq h(\omega) \ a.s.\,[P].$$

The expected value of an integrably bounded random set, denoted by $E(X)$, can be defined by means of Aumann's integral [1].

A mapping $X : \Omega \to \mathcal{F}_{cc}(\mathrm{R}^p)$ is said to be a *fuzzy random variable* if it is Borel-measurable when we consider the d_∞ metric on $\mathcal{F}_{cc}(\mathrm{R}^p)$. A fuzzy random variable is said to be *integrably bounded* if the mapping $X_0 : \Omega \to \mathcal{K}_c(\mathrm{R}^p)$, with $X_0(\omega) = (X(\omega))_0$ for all $\omega \in \Omega$ is an integrably bounded random set.

Puri and Ralescu [7] established the concept of expected value for these integrably bounded fuzzy random variables, as the unique fuzzy set such that $(E(X))_\alpha = E(X_\alpha)$, for every $\alpha \in [0, 1]$.

3 MAIN RESULT

Here, we prove our main result, an theorem about the exchanging of differential and expectation for fuzzy random variables. The result is obtained after embedding the fuzzy random variable by means of its support function, so an exchange result with support function is also obtained.

THEOREM 3.1 *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $T \subseteq \mathrm{R}^k$ be a non-empty, convex and open subset. Let $X : \Omega \times T \to \mathcal{F}_{cc}(\mathrm{R}^p)$ be a mapping satisfying that*

i) for all $t \in T$, the mapping $X_t : \Omega \to \mathcal{F}_{cc}(\mathrm{R}^p)$ with $X_t(\omega) = X(\omega, t)$, is an integrably bounded fuzzy random variable;

ii) the support function associated with the mapping $X_\omega : T \to \mathcal{F}_{cc}(\mathrm{R}^p)$, where $X_\omega(t) = X(\omega, t)$, is Fréchet differentiable at a neighborhood N of $t_0 \in T$, a.s. $[P]$, for every $(\alpha, r) \in [0, 1] \times S^{p-1}$;

iii) there exists a real-valued function $h : \Omega \to \mathrm{R}$, with $h \in \mathrm{L}^1(\Omega, \mathcal{A}, \mathrm{P})$, such that

$$\left\| \frac{\partial}{\partial t} s_{X(\omega,t)}(\alpha, r) \right\| \leq h(\omega) \ a.s.\,[P],$$

for all $t \in N$, and for every $(\alpha, r) \in [0, 1] \times S^{p-1}$.

Then, the following equalities hold:

$$\frac{\partial}{\partial t} s_{EX(\omega,t_0)}(\alpha, r) = \frac{\partial}{\partial t} E(s_{X(\omega,t_0)}(\alpha, r)) \qquad (1)$$

$$\frac{\partial}{\partial t} E(s_{X(\omega,t_0)}(\alpha, r)) = E\left(\frac{\partial}{\partial t} s_{X(\omega,t_0)}(\alpha, r) \right) \qquad (2)$$

for all $(\alpha, r) \in [0, 1] \times S^{p-1}$.

Acknowledgements

This work has been partially supported by a Grant from **Fundación Banco Herrero** and the Grant from the Spanish MEC **DGES PB95-1049**. Authors wish to thank this financial support.

References

[1] Aumann, R.J. (1965). Integrals of set-valued functions. *J. Math. Anal. Appl.* **12**, 1-12.

[2] Dubois, D. and Prade, H. (1982). Towards fuzzy differential calculus. Part 3: Differentiation, *Fuzzy Sets and Systems* **8**, 225-233.

[3] Goetschel, R. and Voxman, W. (1986). Elementary fuzzy calculus, *Fuzzy Sets and Systems* **18**, 31-43.

[4] Kaleva, O. (1987). Fuzzy differential equations, *Fuzzy Sets and Systems* **24**, 301-317.

[5] Puri, M.L. and Ralescu, D.A. (1983). Differentials of fuzzy functions, *J. Math. Anal. Appl.* **91**, 552-558.

[6] Puri, M.L. and Ralescu, D.A. (1985). The concept of normality for fuzzy random variables. *Ann. Prob.* **13**, 1373-1379.

[7] Puri, M.L. and Ralescu, D.A. (1986). Fuzzy random variables. *J. Math. Anal. Appl.* **114**, 409-422.

[8] Rojas-Medar, M., Bassanezi, R.C. and Román-Flores, H. (1999). A generalization of the Minkowski embedding theorem and applications. *Fuzzy Sets and Systems* **102**, 263-269.

[9] Zadeh, L.A. (1975). The concept of a linguistic variable and its application to approximate reasoning. Parts 1, 2, and 3. *Inform. Sci.* **8**, 199-249; **8**, 301-357; **9**, 43-80.

THE CARDINALITY OF A FUZZY SUBSET AS A PART OF THE CARDINALITY OF THE WHOLE FUZZY SET

Pedro Canet Estelrich
Departament de Matemàtiques i Informàtica
Universitat de les Illes Balears
Crra.Valldemossa Km. 7'5
07071 Palma de Mallorca, Spain
Email:pcedmi0@ps.uib.es

Jaume Casasnovas Casasnovas
Departament de Matemàtiques i Informàtica
Universitat de les Illes Balears
Crra.Valldemossa Km. 7'5
Email: jccdmi0@ps.uib.es
07071 Palma de Mallorca, Spain

Abstract

In this paper we propose a definition of the cardinal of a finite fuzzy subset in an analogous situation to the crisp one where we know the relationship between the quantities of elements belonging to the subset and to the whole set.
Keywords: Cardinality of fuzzy sets , fuzzy algebra, natural language

1. INTRODUCTION:

There are solid motivations for undertaking the problem of cardinality of fuzzy sets. On the one hand, we mean a generalisation of the classical cardinality theory for sets. On the other hand, let us mention applications to areas lying far beyond the pure mathematics, such as the meaning of imprecise quantifiers in natural language statements, aggregations of imprecise predicates, decision-making in a fuzzy environment. For instance, we mean the problem of satisfactory answers in databases to queries of the form "How many x's are **p**". Moreover, we need an ordering relation and arithmetic enabling us to compare cardinalities of fuzzy sets and to calculate with them in order to answer to queries of the form: Are there more x's which are p than x's which are q?
In the subject literature, several approaches to cardinality of fuzzy sets can be found (see Wygralak [4,5], Zadeh[6]) For finite fuzzy sets, in [5] we can found the definitions and the properties of the cardinalities as generalised natural numbers, even their arithmetic and comparisons in the same sense which we will use in this paper, where we work cardinality in such form that it allows us to quantify concepts as "the subset B have a half elements of A", which appears in natural language where is assumed that a "twenty per cent of young people are very tall"

2. PRELIMINARY REMARKS:

Each fuzzy set in a universe **M** is characterised by a function $A: \mathbf{M} \to [0\ 1]$, which is called the membership function of that fuzzy set . As usual, the set "suppA" = $=\{x \in \mathbf{M}: A(x) \neq 0\}$. The fuzzy sets refered in this paper are *finite* , i.e. "supp A" is a finite crisp set. If C is a crisp finite set, then |C| denote the "cardinal" of C, the "number of elements which belong to C". $B \subset A$ iff $B(x) \leq A(x)$, $\forall x \in \mathbf{M}$.

Besides, we will use the following definitions:

$$A_t = \{x \in \mathbf{M}: A(x) \geq t\} \ ; t \in [0\ 1] \quad (1)$$

$$[A]_i = \sup \{t \in [0\ 1]: |A_t| \geq i\} \ ; \ i \in N=\{0,1,2,\ldots\} \quad (2)$$

$$m = |A_1| \ ; \ n = |\text{supp } A| \quad (3)$$

Lemma 1: Let A, B finite fuzzy set. The following properties hold true:

a) $[A]_i$ is nonincreasing with respect to i
b) $[A]_i = 1$ for each $i \leq m$
c) $[A]_i = 0$ for each $i > n$
d) $0 < [A]_i < 1$ for each $m < i \leq n$
e) If $m < i \leq n$, then $[A]_i$ is the ith element of the nonincreasingly ordered sequence of all positive values A(x), including their possible repetitions
f) $B \subset A$ implies $[B]_i \leq [A]_i$ for each $i \in N$

Definition 1: Let A be a fuzzy set, the "cardinality" of A is defined (Wygralak [4,5]) as a fuzzy set |A| in N as follows:

$$|A|(i) = \min \{[A]_i , 1 - [A]_{i+1}\} \quad (4)$$

Let us define: $z_A = \min\{i \in N: [A]_i + [A]_{i+1} \leq 1\}$ (5)

Remark 1: By definition:
$|A|(i) = 1 - [A]_{i+1}$ if $i < z_A$, else $[A]_i$

Remark 2 :

a) $[A]_i = 1 - |A|(i - 1)$, for $i \leq z_A$
b) $[A]_i = |A|(i)$, for $i \geq z_A$

Proof: immediate from the definitions

Remark 3 : as a consequence of the above remark :

$$1 - |A|(z_A - 1) = |A|(z_A)$$

Lemma 2 :

a) $|A|(i)$ is no decreasing , for $i = 0,1,\ldots z_A\text{-}1$
b) $|A|(i)$ is nonicreasing for $i = z_A, z_A+1,\ldots n$
c) if $i < z_A$,then $[A]_i > 0'5$ and $[A]_i \leq 0'5$ if $i > z_A$
d) if $i < z_A - 1$, then $|A|(i) < 0'5$, and,
if $i > z_A$, then $|A|(i) \leq 0'5$
e) if $|A|(z_A) < 0'5$ then:
$\max\{|A|(i), i=0,1\ldots n\} = |A|(z_A-1) \geq 0'5$
f) if $|A|(z_A) > 0'5$, then:
$\max\{|A|(i), i=0,1\ldots n\} = |A|(z_A) > 0'5$
g) if $|A|(z_A) = 0'5$, then:
$\max\{|A|(i), i=0,1\ldots n\} = |A|(z_A) = |A|(z_A-1) = 0'5$

Lemma 3 : let A,B be finite fuzzy sets. The following two conditions are equivalent:

a) $\forall i \in N$: $\min\{[A]_i, 1 - [A]_{i+1}\} = \min\{[B]_i, 1 - [B]_{i+1}\}$
b) $\forall i \in N$: $[A]_i = [B]_i$

Definition 2 : Equipotent fuzzy sets

$$A \sim B \text{ iff } |A|(i) = |B|(i)\ \forall i \in N \qquad (6)$$

$$A \sim^s B \text{ iff } |A|(i) - |B|(i) \leq s\ \forall i \in N \qquad (7)$$

Theorem 1 (Wygralak[5]) . Let A,B finite fuzzy sets. The condition A~B is equivalent to each of the following conditions:

a) $\forall i \in N$: $[A]_i = [B]_i$
b) $\forall t \in (0\ 1]$: $|A_t| = |B_t|$

Remark 4 : If $B \subset A$, then the inequality :

$$|A|(i) \leq |B|(i)\ \forall i \in N$$

does not fulfil, but:

a) $|A|(i) \leq |B|(i)$ if $i < \min\{z_B, z_A\}$
b) $|B|(i) \leq |A|(i)$) if $i > \max\{z_B, z_A\}$

3. THE CARDINALITY OF A FUZZY SUBSET AS A PART OF THE THE CARDINALITY OF THE WHOLE FUZZY SET

Definition 3 : Let A be a fuzzy set in **M**. If $0 < p \leq 1$ is a rational number such that $1/p \in N$ and B is a fuzzy subset of A, then:

$$|B| = p|A| \text{ iff : } |B|(i) = |A|(i/p) \quad \text{for each } i \in N \qquad (8)$$

For instance: for p = 1/2

i	$[A]_i$	$\|A\|(i)$	$\|B\|(i)$	$[B]_i$
0	1	0	0	1
1	1	0'1	0'2	1
2	0'9	0'2	0'4	0'8
3	0'8	0'2	0'6	0'6
4	0'8	0'4	0'3	0'3
5	0'6	0'4	0'2	0'2
6	0'6	0'6	0	0
7	0'4	0'4	0	0
8	0'3	0'3	0	0
9	0'3	0'3	0	0
10	0'2	0'2	0	0
11	0'1	0'1	0	0
12	0	0	0	0

where p= ½ ; $z_A = 6$; $z_B = 3$, $|suppA| = 11$, $|suppB| = 5$

Remark 5 : From a fuzzy set , A, with cardinality $|A|(j)$, $j = 0,1,\ldots n$, the cardinality of the fuzzy set, B, $|B|(i) = |A|(i/p)$ results determined with the values of $|A|(j)$, $j = 0, 1/p, 2/p, \ldots$ (0,2,4,6,... in the above example) and this determination does not change depending of the values of the cardinality $|A|(j)$, $i/p < j < (i+1)/p$ (1,3,5,6,7...in the above example). Moreover, the values of $[B]_i$, $i = 0,1,2,\ldots$ are determined by the values of $|B|(i)$.

Remark 6 : The following situation is possible:

for p = 1/2

i	$[A]_i$	$\|A\|(i)$	$\|B\|(i)$	$[B]_i$
0	1	0	0	1
1	1	0'1	0'2	1
2	0'9	0'2	0'3	0'8
3	0'8	0'2	0'6	0'7
4	0'8	0'3	0'3	0'3
5	0'7	0'4	0'2	0'2
6	0'6	0'6	0	0
7	0'4	0'4	0	0
8	0'3	0'3	0	0
9	0'3	0'3	0	0
10	0'2	0'2	0	0
11	0'1	0'1	0	0
12	0	0	0	0

where : $z_A = 6$; $|A|(6) = |B|(3) = 0'6$;
$|A|(4) = |B|(2) = 0'3$;
$z_B = 3$, but $[B]_3 \neq |B|(3)$, in contradiction with Remark 2

Proposition 1: If $z_B / p = z_A$, it is possible to define exactly the cardinality of B from the cardinality of A , iff

$$[A]_{(zB-1)/p+1} = [A]_{zB/p},$$

Proof:

$[A]_{zB/p} = |A|(z_B/p) = |B|(z_B) = 1 - |B|(z_B - 1) =$
$= 1 - |A|((z_B - 1)/p) = [A]_{(zB-1)/p+1}.$

Remark 7: if the condition of the remark 6 does not fulfil, then:

$$A \sim^{s} B \text{ , with } s = [A]_{zA-1/p+1} - [A]_{zA}$$

Proposition 2: If $|B| = p|A|$ and $i_M = \max\{i : i/p \le z_A\}$, then:

$$i_M \le z_B \le i_M + 2$$

Proof:

1. Let suppose that $z_B < i_M$, then we have:

$z_B/p < z_A$; $z_B/p + 1 < (z_B + 1)/p \le z_A$;
$(z_B-1)/p < z_A$; $z_B/p + 1 \le z_A - 1$

a)
$|A|((z_B-1)/p) \le |A|(z_B/p)$, by lemma 2.a,
So, $|B|(z_B-1) \le |B|(z_B)$
And, $1 - [B]_{zB} \le [B]_{zB}$, and $[B]_{zB} \ge 0'5$, by lemma2.e,f,g

b)
$|B|(z_B) \ge |B|(z_B+1)$, by lemma 2.b
So, $|A|(z_B/p) \ge |A|((z_B+1)/p)$

c)
$(z_B + 1)/p < z_A$, because if $(z_B+1)/p = z_A$, then :
$|A|(z_{B/p}) = 1-[A]_{zB/p+1}$,
$|A|((z_B+1)/p) = |A|(z_A) = [A]_{zA} = [A]_{(zB+1)/p}$
and : $1-[A]_{zB/p+1} \ge [A]_{(zB+1)/p}$, by the above b),
So:
$[A]_{za-1} + [A]_{zA} \le [A]_{zB/p+1} + [A]_{(zB+1)/p} \le$
$\le [A]_{zB/p+1} + 1-[A]_{zB/p+1} = 1$
in contradiction with the definition of z_A (Def.1)

So, $(z_B+1)/p < z_A$, and by lemma 2.a:
$|A|(z_B/p) \le |A|((z_B+1)/p)$

c) from b) and c), it results:
$|A|(z_B/p) = |A|((z_B+1)/p)$, and so:
$|B|(z_B) = |B|(z_B+1)$
but $|B|(z_B+1) \le 0'5$, by lemma 2.c.
So: $|B|(z_B) = [B]_{zB} \le 0'5$

e) from a) and d) , it results :
$[B]_{zB} = 0'5$
and $[B]_{zB+1} = [B]_{zB} = |B|(z_B+1) = |A|((z_B+1)/p = 0'5$
and : $|A|(z_B/p) = |B|(z_B) = 0'5$

$1- [A]_{zB/p+1} = 1- [A]_{(zB+1)/p+1} = 0'5$
but $z_B/p + 1 \le z_A - 1$, and $(z_B+1)/p < z_A$ by c)
in contradiction with de definition of z_A

2. Let now suppose that : $z_B > i_M + 2$, then:
$z_B/p > (z_B-1)/p > (z_B-2)/p > i_M/p > z_A$

$|B|(z_B) = |A|(z_B/p) = [A]_{zB/p} \le [A]_{(zB-1)/p} = |A|((z_B-1)/p) =$
$= |B|(z_B-1)$.
Analogously : $[A]_{(zB-1)/p} = |A|((z_B-1)/p)_p \le [A]_{(zB-2)/p} =$
$= |A|((z_B-2)/p) = = |B|(z_B-2)$,
and: $|B|(z_B) \le |B|(z_B-1) \le |B|(z_B-2)$

From where : $1-[B]_{zB-1} \ge 1-[B]_{zB} \ge [B]_{zB}$
But : $[B]_{zB-1} \ge [B]_{zB}$,, by lemma 1.a, and we can assume:
$[B]_{zB-1} = [B]_{zB}$
On the other hand $[B]_{zB-1} > 0'5$, by lemma 2.c, and we also can assume $[B]_{zB} > 0'5$ and $1-[B]_{zB} < [B]_{zB}$, which results contradictory

Remark 8 : it is possible that $z_B = i_M$, i_M+1 or i_M+2
For instance:

i	[B](i)	|B|(i)	|A|(i)	[A](i)
0	1	0'1	0'1	1
1	0'9	0'3	0'2	0'9
2	0'7	0'5	0'3	0'8
3	0'5	0'5	0'5	0'7
4	0'4	0'4	0'5	0'5
5	0'3	0'3	0'5	0'5
6	0	0	0'5	0'5
7	0	0	0'45	0'45
8	0	0	0'4	0'4
9	0	0	0'35	0'35
10	0	0	0'3	0'3
11	0	0	0'2	0'2
12	0	0	0	0

where p= ½ ; $z_A = 4$; $i_M = 2$; $z_B = 3$, |suppA|=11 ,
|suppB|=5
and even it is possible to find examples where $z_B = i_M + 2$ and $pi_M + 1 < z_A$, for instance:

i	[B](i)	|B|(i)	|A|(i)	[A](i)
0	1	0'1	0'1	1
1	0'9	0'4	0'2	0'9
2	0'6	0'5	0'3	0'8
3	0'5	0'5	0'4	0'7
4	0'4	0'4	0'5	0'6
5	0'3	0'3	0'5	0'5
6	0	0	0'5	0'5
7	0	0	0'5	0'5
8	0	0	0'5	0'5
9	0	0	0'5	0'5
10	0	0	0'4	0'4
11	0	0	0'4	0'4
12	0	0	0'4	0'4
13	0	0	0'3	0'3
14	0	0	0'3	0'3
15	0	0	0'3	0'3
16	0	0	0	0

where p = 1/3 ; $z_A = 5$; $i_M = 1$; $z_B = 3$, |suppA|=15,
|suppB|=5

Proposition 3: If $|B| = p|A|$,

a) $[B]_{i+1} = [A]_{i/p+1}$ for each $i \in N$, such that: $i < i_M$

b) $[B]_i = [A]_{i/p}$, for each $i \in N$, such that: $i > z_B$

Proof:

a) if $i < i_M$, then $i < z_B$, by Prop.2, and $i/p < z_A$, by the definition of i_M.

Therefore: $1 - [B]_{i+1} = |B|(i) = |A|(i/p) = 1 - [A]_{i/p+1}$

b) if $i > z_B$, then $i > i_B$, and $i/p > z_A$

Therefore : $[B]_i = |B|(i) = |A|(i/p) = [A]_{i/p}$

Remark 9 : In order to investigate the relationship between $|A_t|$ and $|B_t|$, for each $t \in [0\ 1]$, we mean that, after the definitions of the Preliminary Remarks, is easy to see $|A_t| = \max\{j : [A]_j \geq t\}$. So, the relationship between $|A_t|$ and $|B_t|$ is a consequence of the relationship between $[B]_i$ and $[A]_{i/p}$, i= 0,1,2,..., which has been proved in the above proposition.

We cannot establish an only relation for each $t \in [0\ 1]$, but it depends of the value of t.

So, we can prove propositions as the following one:

Proposition 4: If $|B| = p|A|$ and $|A_t| \leq z_A$ (which occurs when $t > 0'5$, by lemma 2.c), then:

$$|A_t| - 1/p|B_t| \leq 1$$

Proof:

For a major simplicity , let $|A_t| = \max\{j : [A]_j \geq t\} = j_M$

If $j_M \leq z_A$, then $j_M - 1 < z_A$;

If $k_M = \max\{k \in N: i/p \leq (|A_t|-1)\} =$

$= \max\{i \in N: i/p \leq (j_M - 1)\}$, then: $k_M < i_M$ (defined in Prop.2), and $k_M/p + 1 \leq j_M - 1 + 1 = j_M$

So, by Prop.3.a: $[B]_{kM+1} = [A]_{kM/p+1}$

And $[A]_{kM/p+1} \geq [A]_{jM} \geq t$, by lemma 1.a

but $[B]_{kM+2} = [A]_{(kM+1)/p+1}$ or $[B]_{kM+2} = [A]_{(kM+2)/p}$ and in every case $[B]_{kM+2} < t$, because $(k_M+1)/p + 1 > j_M - 1 + 1 = j_M$, and $(k_M + 2)/p = (k_M + 1)/p + 1/p > j_M - 1 + 1 = j_M$

So, $|B_t| = \max\{i : [B]_i \geq t\} = k_M + 1$, and we can write:

$1/p \geq (j_M - 1) - k_M/p = (j_M - 1) - 1/p(k_M + 1 - 1) =$

$= (|A_t| - 1) - 1/p(|B_t| - 1) \leq 1/p$

and operating : $|A_t| - 1/p|B_t| \leq 1$

Remark 10: Analogously, if $|A_t| > z_A$ (which occurs when $t \leq 0'5$, lemma 2.c) , $|A_t| - 1/p|B_t| \leq 1/p$

Proposition 5 : If $|B| = p|A|$:

$$|SuppA| - 1/p|SuppB| < 1/p$$

Proof:

Like A and B are finite fuzzy sets, there are:

$|suppB| = \max\{i : [B]_i > 0\} = i_m \geq z_B$

$|suppA| = \max\{j : [A]_j > 0\} = j_m \geq z_A$

but: $[B]_{im} = |B|(i_m) = |A|(i_m/p) = [A]_{im/p}$

and $j_m - i_m/p < 1/p$, because if $j_m - i_m/p \geq 1/p$, we would have: $j_m \geq (i_m/p + 1)/p$, and:

$[B]_{im+1} = |B|(i_m+1) = |A|((i_m+1)/p) = [A]_{(im+1)/p} \geq [A]_j > 0$

which contradicts the definition of i_m .

5. CONCLUSIONS:

The proposed definition of the cardinal of a fuzzy subset allows us to work with the concept of a rational part of the whole set or universe. The relationship between the corresponding values that characterize the cardinal of a fuzzy set in (Wygralak [4]) , has been investigated.

Acknowledgements

We thank the referee for valuable comments

References

[1] J.Casasnovas. El método de acumulación aplicado al razonamiento numérico.Actas del VIII Congreso español sobre tecnologias y lógica fuzzy.419-424. Pamplona,Spain.Septiembre1998

[2] D.Ralescu.*Cardinality, quantifiers, and the aggregation of fuzzy criteria*. Fuzzy Sets and Systems 69(1995)355-365

[3] E.Trillas,C.Alsina,JMTerricabras: *Introducción a la Lógica Borrosa.Ariel Matemàtica*. Barcelona 1995

[4] M.Wygralak.*Vaguely Defined Objects.Representations.Fuzzy Sets, and No Classical Cardinality Theory*.Kluwer,Dordrecht,Boston,London,1996

[5] M.Wygralak *Questions of cardinality of finite fuzzy sets*.Fuzzy sets and Systems 102(1999)185-210

[6]L.A.Zadeh.*A computational approach to fuzzy quantifiers in natural languages*.Comput.Math.Appl.9(1983)149-184

CUTS, SLICES AND COMPUTABILITY.

R. Morales Bueno, J.L. Pérez de la Cruz, F. Triguero, R. Conejo
Dept. Lenguajes y Ciencias de la Computación. E.T.S.I. Informática.
Universidad de Málaga. Campus Teatinos 29071. Málaga. España.
e-mail: perez@lcc.uma.es

Abstract

In this paper several classes of fuzzy functions are studied and related to the computations of a fuzzy Turing machine. These classes are put into relation with the ones previously proposed in the literature.

Key words: fuzzy function, recursive function, computability, Turing machine.

1. INTRODUCTION

The computability of fuzzy functions has been the object of some attention since the first paper by E. Santos [8]; we must mention especially the work of Gerla [1,2]. Two main approaches have been followed to the concept: first, there are definitions of fuzzy computing devices or programs [8,9] [6,7]; on the other hand, there are definitions in terms of the recursiveness of the associated crisp functions [2,3]. Of course, both approaches converge.

In the present paper, we summarize work previously done and provide a new framework where these concepts fit. First, we prove that there are only three classes of functions definable in terms of recursiveness of cuts and slices (section 2) and also consider the bounded case (ibid.) These are new results. Then, in section 3 we will summarize definitions concerning fuzzy Turing machines and fuzzy computations. Some of these concepts have been presented in [5]. Finally, in section 4 the correspondence between both approaches is proven.

2. FROM α-CUTS TO RECURSIVENESS

In the following $W = \{0, \alpha_1, ..., \alpha_n\}$, $0 < \alpha_1 < ... < \alpha_n = 1$, will be an ordered finite semiring whose operations are $\oplus$ and $\otimes$ and whose order is $\leq$. The top of W will be the unit of $\otimes$, noted as 1. Anagolously, the bottom of W will be the zero of $\oplus$, noted as 0.

A fuzzy subset (W-set) of A is a (crisp) total function from A to W. We will note W-sets of A by greek lowercase letters subindexed by A. For example, μ_A, λ_A are W-sets of A. The α-cut of μ_A is the crisp set $\{x \in A \mid \mu_A(x) \geq \alpha\}$. The α–slice of μ_A is the crisp set $\{x \in A \mid \mu_A(x) = \alpha\}$.

A W-function from A to B is a W-set of $A \times B$. We will use greek lowercase letters subindexed by $A \times B$ to represent W-functions from A to B. The value $\phi_{A\times B}(x, y)$ can be interpreted as the degree with which the function $\phi_{A\times B}$ yields y when its argument is x. We will omit the subindex $A \times B$ when it is clear from the context. The image of $x \in A$ by a W-function $\phi_{A\times B}$ is the W-set of B $\zeta_{\phi,x}$ given by $\zeta_{\phi, x}(y) = \phi_{A\times B}(x, y)$.

We will only consider functions whose domain A and range B are effectively enumerable, i.e., there are computable bijections between N and A and between N and B. Therefore, A and B can be considered ordered by these bijections.

Definition 1. A W-function ϕ is partially recursive iff every α-cut ($\alpha \neq 0$) of ϕ is recursively enumerable.

Comment: Gerla [2] uses the denomination "partial recursive function". By restricting ϕ to its support, ϕ could be viewed as a partial function from $A \times B$ to W and hence this denomination makes sense. However, since ϕ is in fact total, we prefer the denomination "partially recursive function".

Definition 2. A W-function ϕ is pseudorecursive iff every α–slice ($\alpha \neq 0$) of ϕ is recursively enumerable.

Definition 3. A W-function ϕ is totally recursive iff every α-cut ($\alpha \neq 0$) of ϕ is recursive.

Comment: Gerla [2] uses the denomination "function with recursive graph". Our proposal aims to show the similarity with the remaining classes of recursive functions.

The proviso $\alpha \neq 0$ could be omitted in definitions 1 and 3. In fact, the 0-cut is the product $A \times B$, hence it is recursive. On the contrary, since the complementary of a recursively enumerable set may be not recursively enumerable, the restriction plays a substantial rule in definition 2.

Proposition 1. i) If ϕ is pseudorecursive, then ϕ is partially recursive; ii) if every α–slice of ϕ ($\alpha \neq 0$) is recursive, then ϕ is totally recursive.

Proof: i) the α_i-cut of ϕ ($\alpha \neq 0$) can be expressed by the union of the α_i-slice, α_{i+1}-slice, ... Therefore, if the α–slices ($\alpha \neq 0$) are recursively enumerable then the α-cuts ($\alpha \neq 0$) are recursively enumerable, too. The same argument -replacing "recursively enumerable" with "recursive"- proves ii). In fact, the finite union of recursively enumerable sets is recursively enumerable. The same assertion holds for recursive sets.

Proposition 2. If ϕ is totally recursive, then every α–slice of ϕ is recursive.

Proof: the α_i-slice of ϕ can be expressed by the difference between the α_i-cut and the α_{i+1}-cut. Therefore, if the α-cuts are recursive, then the α–slices are recursive, too.

Comment. These definitions and propositions can be summarized as follows: by combining some of the above conditions, namely that every α–slice/cut (including/excluding $\alpha=0$) is recursive/recursively enumerable, only three classes of *W*-functions arise: those of partially recursive, pseudorecursive and totally recursive functions. Moreover, if ϕ is totally recursive then ϕ is pseudorecursive and if ϕ is pseudorecursive then ϕ is partially recursive. Table 1 shows these results:

excluding $\alpha=0$	r. e.	α-cuts	partially rec.
excluding $\alpha=0$	r. e.	α–slices	pseudorec.
excluding $\alpha=0$	rec.	α-cuts	totally rec.
excluding $\alpha=0$	rec.	α–slices	totally rec.
including $\alpha=0$	r. e.	α-cuts	partially rec.
including $\alpha=0$	r. e.	α–slices	totally rec.
including $\alpha=0$	rec.	α-cuts	totally rec.
including $\alpha=0$	rec.	α–slices	totally rec.

Table 1.

Definition 4. A set $D \subseteq A \times B$ is recursively bounded iff

i) D is recursive.

ii) there exists a recursive function $f:A \to B$ such that for every $x \in A$ $\{(x, y) \in D\} \subseteq \{(x, y) \mid y \leq f(x)\}$.

Definition 5. A *W*-function ϕ is bounded recursive iff every α-cut ($\alpha \neq 0$) of ϕ is recursively bounded.

Comment: since the 0-cut of ϕ is $A \times B$ and hence infinite, the proviso $\alpha \neq 0$ cannot be omitted in this definition.

Proposition 3. If ϕ is bounded recursive, then every α–slice of ϕ ($\alpha \neq 0$) is recursively bounded.

Proof: the α_i-slice of ϕ can be expressed by the difference between the α_i-cut and the α_{i+1}-cut. Therefore, if the α-cuts ($\alpha \neq 0$) are recursive, then the α–slices ($\alpha \neq 0$) are recursive, too. On the other hand, let f_1, ..., f_n be the bounding recursive functions of the α_1, ..., α_n -cuts. Obviously, f_1 bounds every α–slice ($\alpha \neq 0$), hence every α–slice ($\alpha \neq 0$) is recursively bounded.

3. FUZZY COMPUTATIONS

In the following we summarize and arrange some concepts presented in previous papers [4] [5].

Definition 6.- A *W*-Turing machine (*W*-TM) is a 4-uple $Z=(V, S, p, s_0)$ where V and S are finite nonempty sets, $V \cap S = \emptyset$; $s_0 \in S$ and p is a *W*-function from $S \times V$ to $(V \cup \{L, R\}) \times (S \cup \{h\})$; $L, R \notin V$, $h \notin S$.

In the definition above S is the *state set* , s_0 the *initial state*, V the set of *simbols*. Moreover, p is the *transition W-function*, namely $p(s, c, z, s')$ represents the degree of validity of the choice of the act (z, s') at the state s and when the tape symbol c is scanned. Such an act is to move to the rigth (if z is R), to the left (if z is L) or to replace c by z (if $z \in V$) and successively to asume the state s' (stop if s' is h). We shall assume that in V there is a special symbol # that stands for blank.

For every $u \neq 0$, we can write $(s, c) \to^u (z, s')$ to denote that $p(s, c, z, s')=u$ and $(s,c) \to (z,s')$ to denote that $p(s, c, z, s')=1$. The *W*-function p is characterized by a finite set of arrows of the type above.

Definition 7. An *elementary description* is an element of $(\{\varepsilon\} \cup (V-\{\#\})V^*)$ $(S \cup \{h\})V$ $(V^*(V-\{\#\}) \cup \{\varepsilon\})$, i. e., a string of the form *usaw*, where s is a state (or the halting state h), u is the empty string or a string over V whose first symbol is not #, a is a symbol of V, and w is the empty string or a string over V whose last symbol is not #. An elementary description expresses the present state (s), the content of the tape (*uaw*) and the position and content of the scanned square (a). Elementary descriptions will be denoted by d, d_1, d_2, ... The set of all elementary descriptions of T will be denoted $\mathcal{D}(T)$. We will omit the argument T when no ambiguity can arise.

Definition 8. Let $d \in \mathcal{D}$ be an elementary description and let p be the transition function of T. The set of successors of d, σ^d, is a *W*-set of $\mathcal{D}$ defined as follows:

i) if $d=uscw$, $d'=us'c'w$ then $\sigma^d(d')=p(s,c,c',s')$.

ii) if $d=uscc'w$, $d'=ucs'c'w$, then $\sigma^d(d')=p(s,c,R,s')$.

iii) if $d=usc$, d'=ucs'#, then $\sigma^d(d')=p(s,c,R,s')$.

iv) if $d=uc'scw$, $d'=us'c'cw$, then $\sigma^d(d')=p(s,c,L,s')$.

v) if $d=scw$, $d'=s'\#cw$, then $\sigma^d(d')=p(s,c,L,s')$.

vi) otherwise $\sigma^d(d')=0$.

Definition 9. A description $\delta_{\mathcal{D}}$ is a W-set of $\mathcal{D}$. $\delta_{\mathcal{D}}$ is an initial description iff $\delta_{\mathcal{D}}= \{1/s_0\#\}$.

Definition 10.- An elementary description $d=usaw$ is *final* iff $s=h$. We will call u the output string of d. A description $\delta_{\mathcal{D}}$ is final iff for every $d\in\mathcal{D}$ such that $\delta_{\mathcal{D}}(d)\neq 0$, d is an elementary final description.

Let us define now the transitions between fuzzy descriptions. Notice that since every elementary description yields a finite number of elementary descriptions in one step, a fuzzy description with finite support must yield another finite description.

Definition 11.- We say that T generates δ' from δ in one step ($\delta \Rightarrow_T \delta'$) iff δ is not a final description and δ' is the fuzzy subset of $\mathcal{D}$ given by $\delta'(d') = \sum_{d\in\mathcal{D}} \delta(d) \otimes \sigma^d(d')$.

Definition 12.- Let T be a W-TM. The sequence generated by T with input x is $\{\delta_i(x)\}$ defined as follows: a) $\delta_0(x) = \{1/s_0\, x\}$; b) $\delta_i(x) \Rightarrow_T \delta_{i+1}(x)$.

For every description δ_i we can define a W-set μ which can be thought as already computed by T. Let us call it the "output set" of δ_i. Formally:

Definition 13.- Let δ be a description. Let $d=uhaw$ be a generic final elementary description. The output set of δ is the W-set of V^* μ given by $\mu(u)= \sum_{a,w} \delta(uhaw)$. If the sequence generated by T with input x is $\{\delta_i(x)\}$, then the output sequence of T with input x is $\{\mu_i(x)\}$ where every $\mu_i(x)$ is the output set of $\delta_i(x)$.

Notice that these sequences may be finite or infinite and always are fuzzy chains (nondecreasing w. r. t. fuzzy set inclusion). Since W is a finite lattice, the set of W-sets of V^* is a complete lattice and there is always a lub of $\{\mu_i(x)\}$ given by $\mu_x = \cup\, \mu_i(x)$.

Definition 14.- Let T be a W-TM and μ a W-set of V*. T computes μ with input x iff $\mu = \mu_x$. Let ϕ be a W-function from U^* to V^*. If for every $x\in U^*$, T computes the image $\zeta_{\phi,x}$ we say that T computes ϕ and ϕ is computable. Note: In [5] the same concepts were phrased as "T politropically computes ϕ and ϕ is politropically computable".

Definition 15.- Let T be a W-TM and μ a W-set of V^*. T computes monotropically μ with input x iff i) $\mu = \mu_x$; ii) for every step i and every element y of V^*, if $\mu_i(x)(y)\neq 0$ then $\mu_i(x)(y) = \mu_x(y)$. Part ii) could be stated as follows: for every element y of V^* the degree $\mu_x(y)$ is computed in a single step.

Let ϕ be a W-function from U^* to V^*. If for every $x\in U^*$, T computes monotropically the image $\zeta_{\phi,x}$ we say that T computes monotropically ϕ and ϕ is monotropically computable. Obviously, if T computes monotropically μ with input x, T computes μ with input x, and if ϕ is monotropically computable, then ϕ is computable.

Definition 16.- Let T be a W-TM and μ a W-set of V^*. T computes monotonically μ with input x iff i) $\mu = \mu_x$; ii) for every step i and every $y\in V^*$, if $\mu_i(x)(y)\neq 0$ then $\mu_i(x)(y') = \mu_x(y')$ for every $y'\in V^*$ such that $y'\leq y$; iii) if the support of μ_x is finite, then $\{\mu_i(x)\}$ is a finite sequence. Part ii) could be stated as follows: for every element y of V^* the degree $\mu_x(y)$ is computed in a single step, and at that step the degree of every $y'< y$ has been already computed. Note: In [5] the third condition was not stated, so affecting the correction of the proposition corresponding to the proposition 5 of present paper.

If for every $x\in U^*$, T computes monotonically the image $\zeta_{\phi,x}$ we say that T computes monotonically ϕ and ϕ is monotonically computable. Obviously, if T computes monotonically μ with input x, T computes monotropically μ with input x, and if ϕ is monotonically computable, then it is monotropically computable.

Definition 17.- Let T be a W-TM and μ a W-set of V^*. Let $\{\mu_i\}$ be the output sequence of T with input x. T computes finitely μ with input x iff $\{\mu_i(x)\}$ is finite and $\mu=\mu_x$.

Let ϕ be a W-function from U^* to V^*. If for every $x\in U^*$, T computes finetely the image $\zeta_{\phi,x}$ we say that T computes finetely ϕ and ϕ is finitely computable. Note: In [5] the same concepts were phrased as "T constructs ϕ and ϕ is constructible".

4. EQUIVALENCES BETWEEN RECURSIVENESS AND COMPUTABILITY.

Lemma 1. Let T be a crisp machine enumerating a sequence $S=\{(u_i, v_j)\}$ of elements of $U^*\times V^*$. Then for every $\alpha\in W$ there exists a W-TM T'_α such that $\mu_{u_i}(v_j) = \alpha$ if there exists a pair in S of the form (u_i, v_j), otherwise $\mu_{u_i}(v_j) = 0$.

Proof (sketch): The computation of T'_α could be described as follows: with input u_i, T'_α simulates the computation of T. Whenever an output (w_i, v_j) is generated, the first component w_i is compared with u_i. If they are equal, T'_α outputs v_j with degree α and goes on simulating T; otherwise it goes on without any output. Notice that this is a monotropical computation.

Proposition 4. The *W*-function ϕ is computable iff it is partially recursive.

Proof: vd. [2] and [5].

Proposition 5. The *W*-function ϕ is monotropically computable iff it is pseudorecursive.

Proof (sketch): Let T be the *W*-TM that computes ϕ. For every $\alpha \neq 0$, let us define a crisp TM T'_α such that T'_α enumerates the α-slice. T'_α sequentially writes the elements in U^* in some order; for each one, it simulates n steps of the computation of T. "Dovetailing" in the elements of U^* and the number of steps, T halts and outputs (u, v) iff v is generated with degree α. This is an effective proceduce to enumerate the α-slice. On the other hand, let us assume that every α-slice is r. e ($\alpha \neq 0$). Then there exist n crisp TM $T_1, ..., T_n$ that enumerate the α-slices. By lemma 1, there also exist n *W*-TM $T'_1, ..., T'_n$ whose parallel composition is able to compute motrotropically the whole output.

Proposition 6. The *W*-function ϕ is monotonically computable iff it is totally recursive.

Proof: vd. [5], proposition 5.

Proposition 7. The *W*-function ϕ is finitely computable iff it is bounded recursive.

Proof (sketch): In [4] it is proven that ϕ is finitely computable iff ϕ is totally recursive and there exists a crisp recursive function $s_\phi(x)$ such that for every x, $s_\phi(x) = sup(\zeta_{\phi,x})$. So, it suffices to show that if ϕ is bounded recursive, then $s_\phi(x)$ is recursive. If ϕ is bounded recursive, then the first α-cut $A_1 = \{(x, y) \mid \phi_{A\times B}(x, y) > 0\}$ is such that there exists a recursive function $f: A \rightarrow B$ such that $A_1 \subseteq \{(x, y) \mid y \leq f(x)\}$. From this f it is possible to compute in a finite number of steps all the possible images of x, hence to count the size of $\zeta_{\phi,x}$, so yielding a recursive $s_\phi(x)$.

Proposition 8. The *W*-function ϕ is computable iff it is *W*-computable (in Santos' sense).

Proof: By proposition 4, ϕ is computable iff it is partially recursive. In [2] it is proven that ϕ is computable (in Santos' sense) iff it is partially recursive.

Proposition 9. The *W*-function ϕ is monotropically computable iff it is strong partial recursive (in Gerla's sense).

Proof: By proposition 5, ϕ is monotropically computable iff it is pseudorecursive. Hence, there exists an effective procedure to set the value of $\phi(x, y)$, assuming that $\phi(x, y) \neq 0$ and $\phi(x, y)$ is computable restricted to $\phi(x, y) \neq 0$. Therefore it is *SPR*. In other words, the three concepts of monotropically computable, pseudorrecursive and strong partial recursive are the same.

5. CONCLUSIONS

An overview of some of the concepts of the theory of fuzzy computability has been presented and developed in this paper. Two threads, namely those of fuzzy Turing machines and recursiveness of cuts and slices, have been followed along the paper. The resulting, unifying view is somehow clear and concise. We think that little more could be done in the case of a finite set *W*.

Perhaps this work could be generalized for the case of a recursive set *W* provided with recursive order and operations. We are currently trying this extension.

REFERENCES

[1] L. Biacino and G. Gerla: Fuzzy subsets: A constructive approach. *Fuzzy Sets and Systems* **45** , 161-168 (1992).

[2] G. Gerla: Turing L-Machines and Recursive Computability for L-Maps. *Studia Logica* **XLVIII**, 179-192 (1989).

[3] L. Harkleroad: Fuzzy Recursion, Ret's and Isols. *Zeitschr. F.. Math. Logik Und Grundlagen D. Math.* **30**, 425-436 (1984).

[4] R. Morales-Bueno, J. L. Perez de la Cruz, B. Clares, and R. Conejo: Estudio de una clase de *W*-funciones calculables. *III Congreso Español Sobre Tecnologías y Lógica Fuzzy, Santiago de Compostela*, 1993, Santiago de Compostela, 1993, p. 153-160.

[5] Morales R., Pérez de la Cruz J. L., Clares B. , and Conejo R.: Acerca de la calculabilidad de las funciones difusas. *Actas Del IV Congreso De La Asociación Española De Tecnologías y Lógica Fuzzy, Blanes (Gerona)*, 1994, editores F. Esteva and P. García, Blanes (Gerona), 1994, p. 115-120.

[6] R. Morales, J. L. Pérez de la Cruz, R. Conejo, and B. Clares: A family of fuzzy programming languages. *Fuzzy Sets and Systems* **87**, 167-179 (1996).

[7] J. L. Pérez de la Cruz, R. Morales-Bueno, R. Conejo, and B. Clares: Caracterización mediante programas de ciertas clases de funciones *W*-calculables. *V Congreso Español Sobre Tecnologías y Lógica Fuzzy, Murcia*, 1995, Murcia, 1995, p. 141-146.

[8] E. S. Santos: Fuzzy algorithms. *Information and Control* **17**, 326-339 (1970).

[9] E. S. Santos, *Fuzzy Automata and Decision Processes*, editors M. M. Gupta, G. N. Saridis, and B. R. Gaines (North-Holland, New York, 1977), p. 133.

ON THE SYMMETRY OF FUZZY SETS

V. Gisin
Finance Academy under the Government of the Russian Federation
Department of Mathematics
Kibalchicha 1, 129848 Moscow, Russia
gisin@glasnet.ru

Summary

In this paper a notion of symmetry of a fuzzy set with a finite support is introduced which takes into account both the intralevel symmetry and the interlevel symmetry. It is shown that the ranked distribution of the relative level cardinalities of a set with the minimal index of symmetry is close to a Pareto distribution. An estimation of the maximal number of fuzzy elements of a fuzzy set with the minimal index of symmetry is given.

Keywords: Fuzzy set, Permutation, Symmetric group, Pareto distribution .

1 INTRODUCTION

In [2] Yu. Shreider gave a philosophical justification of the principle of the minimal symmetry. According to this principle, the minimal symmetry (in a broad sense) is characteristic for natural systems. Applying the principle of the minimal symmetry to fuzzy sets that can be considered as natural (by their origin) we may draw conclusions concerning their membership functions. The concept of symmetry introduced in this paper being combined with the principle of the minimal symmetry allows to give an explanation of the appearance of Pareto distributions in some "fuzzy" situations (for example of the effect of "gross tails" in financial modeling, see [1, chapter 4]).

2 THE INDEX OF SYMMETRY OF A FUZZY SET

Let us consider a fuzzy set with a finite support X and a membership function μ. Let μ_1, μ_2, ..., μ_k be the elements of $\mu(X)$ and let

$$X_i = \{x \,|\, \mu(x) = \mu_i\},$$

$n_i = |X_i|$, and $n = |X|$ ($|\cdot|$ stands for the number of elements). Suppose that the sequence n_i is nonincreasing. A permutation g of X is compatible with the membership function μ iff all sets X_i are invariant with respect to g. So the following group

$$S(X,\mu) = \prod_{i=1}^{k} S(X_i)$$

can be treated as the symmetric group of the fuzzy set (X,μ) (given a set A, by $S(A)$ we denote the symmetric group of A). Clearly

$$|S(X,\mu)| = \prod_{i=1}^{k} n_i!.$$

The group $S(X,\mu)$ takes into account only one type of the symmetry of (X,μ) (its "intralevel" symmetry). We are going to construct the group of the "interlevel" symmetry.

Let X' be a subset of X. We say that $V \subset X'$ is a colevel in X' if V contains exactly one element of each level of X'. In other words the restriction $\mu : V \to \mu(X')$ is one-to-one. By a colevel decomposition of (X,μ) we understand a representation

$$X = \bigcup_{j=1}^{l} V_j \tag{1}$$

such that V_t is a colevel in

$$\bigcup_{j=t}^{l} V_j$$

for every $t = 1, \ldots, l$. We put

$$m_j = |V_j|$$

for $j = 1, \ldots, l$. It can be easily seen that $l = n_1$ and $k = m_1$.

Given a colevel decomposition (1), let $S'(X, \mu)$ be the group of permutations of X which leave every colevel from the decomposition invariant. It is clear that $S'(X, \mu)$ is determined uniquely up to isomorphism. We have

$$S'(X, \mu) = \prod_{j=1}^{l} S(V_j)$$

and

$$|S'(X, \mu)| = \prod_{j=1}^{l} m_j!.$$

The group

$$S(X, \mu) \times S'(X, \mu)$$

seems to be more suitable to evaluate the total symmetry of (X, μ) than the "traditional" symmetric group $S(X, \mu)$.

We say that the number

$$s = \frac{1}{n!} \times \prod_{i=1}^{k} n_i! \times \prod_{j=1}^{l} m_j!.$$

is the *index of symmetry* of (X, μ).

It can be easily seen that $s \leq 1$. Indeed, let us consider the map

$$S(X, \mu) \times S'(X, \mu) \to S(X, \mu)$$

defined by $(a, a') \mapsto a' \circ a$ and show that it is injective. Let $(a, a') \neq (b, b')$. First suppose that $a \neq b$. Then $a(x) \neq b(x)$ for some $x \in X$. So $a(x)$ and $b(x)$ belong to different colevels, and $a'(a(x)) \neq b'(b(x))$ because a' and b' preserve colevels. If $a = b$ but $a' \neq b'$ then there exists $y \in X$ such that $a'(y) \neq b'(y)$. So, for $x = a^{-1}(y)$ we have $a'(a(x)) \neq b'(b(x))$.

We have $s = 1$ if either $k = 1$ or $k = n$. In the following section we discuss situations when s takes its minimal value.

3 FUZZY SETS WITH THE MINIMAL INDEX OF SYMMETRY

To describe fuzzy sets (X, μ) with the minimal value of the index of symmetry we follow ideas from [2].

We say that a pair (i, j) is a Pareto point (with respect to a colevel decomposition (1) of (X, μ)) if

$$X_i \cap V_j \neq \emptyset,$$

$$X_i \cap V_{j+1} = \emptyset, \qquad X_{i+1} \cap V_j = \emptyset.$$

Theorem 1 *Let (X, μ) have the minimal index of symmetry s (with n and $k \geq 4$ fixed). Then the sequence of all Pareto points can be divided into three nonempty parts:*

$$(1, n_1), \quad (2, n_2) \quad , \ldots, (t-1, n_{t-1}), \tag{2}$$

$$(t, n_t) = (m_p, p), \tag{3}$$

$$(m_{p-1}, p-1), \quad (m_{p-2}, p-2) \quad , \ldots, (m_1, 1) \tag{4}$$

such that the following conditions are verified. For Pareto points (i, j), $(i-1, j')$ from (2) we have $n_{i-1} - n_i \geq 2$ and

$$n_{i-1} m_{j'} - m_j \leq n_i m_j \leq n_{i-1} m_{j'} + m_{j'}$$

Analogous inequalities hold for Pareto points from (4).

Example 1 *Let $n = 10$. We have the minimal index of symmetry $s \approx 0.02$ in the following cases:*

(1) $n_1 = 4$, $n_2 = 2$, $n_3 = n_4 = n_5 = n_6 = 1$;

(2) $n_1 = 4$, $n_2 = n_3 = 2$, $n_4 = n_5 = 1$;

(3) $n_1 = 6$, $n_2 = 2$, $n_3 = n_4 = 1$ (the dual of (1));

(4) $n_1 = 5$, $n_2 = 3$, $n_3 = n_4 = 1$ (the dual of (2));

Theorem 2 *The number of levels k such that*

$$k \ln k \approx n$$

provides the minimal value of the index of symmetry.

¿From Theorem 1 it follows that Pareto points are close to a hyperbola $nm = const$ and the distribution of n_i/n can be considered as a discrete Pareto distribution. The proof of Theorem 1 is rather technical. We give its "approximate" version.

Let $y = f(x)$ be a smooth decreasing function approximating $m_j = f(n_i)$ at Pareto points. Using the Stirling formula we present $\ln(n! \cdot s)$ as the sum of the following summands:

$$\sum_{i=1}^{k} (0.5 \ln 2\pi + 0.5 \ln n_i + n_i \ln n_i - n_i)$$

and

$$\sum_{j=1}^{l} (0.5 \ln 2\pi + 0.5 \ln m_j + m_j \ln m_j - m_j).$$

Then, up to a constant (depending on k and n), $\ln s$ is approximately equal to the sum of the following two integrals:

$$\int_1^k (0.5 \ln y + y \ln y)\, dx$$

and

$$\int_1^l (0.5 \ln 2\pi + 0.5 \ln x + x \ln x)\, dy. \qquad (5)$$

The integral in (5) can be presented as

$$-\int_1^k (0.5 \ln 2\pi + 0.5 \ln x + x \ln x) y'\, dx.$$

So we have to minimize

$$\int_1^k \Phi\, dx$$

where

$$\Phi = (0.5 + y) \ln y - 0.5 \ln 2\pi \cdot y' - (0.5 + x) \ln x \cdot y'$$

under the following condition

$$\int_1^k y\, dx = n.$$

We put

$$F(x, y, y') = \Phi + \lambda y.$$

Now applying the Euler equation

$$\frac{\partial F}{\partial y} - \frac{d}{dx} \frac{\partial F}{\partial y'} = 0$$

we get

$$\frac{1}{2x} + \ln x + \frac{1}{2y} + \ln y = const.$$

Therefore

$$xy \cdot \exp\left(\frac{1}{2x}\right) \cdot \exp\left(\frac{1}{2y}\right) = const. \qquad (6)$$

Formula (6) shows that if x and y are not too small then $xy \approx const$, i.e. the relation between x and y is similar to that in a Pareto distribution.

Now assume that the segment $[0; 1]$ is divided into k equal intervals by the range of the membership function μ. If

$$1 = \mu_1 > \mu_2 > \ldots > \mu_k = \frac{1}{k}$$

then the number of fuzzy elements in X, that is

$$\tilde{n} = \sum_{i=1}^{k} \mu_i n_i,$$

is maximal. The following estimation can be obtained

$$\tilde{n} \approx n + \ln k - k.$$

If $n \approx k \ln k$ (as in Theorem 2) then

$$\tilde{n} \approx n \cdot \left(1 + \frac{1}{k} - \frac{1}{\ln k}\right).$$

For example if $n = 100$ then $k = 30$ and $\tilde{n} \approx 74$.

References

[1] A. Shiryaev (1998). The Fundamentals of Stochastic Financial Mathematics. Phasis, Moscow, 1998, 1018 pp.

[2] Yu. Shreider, A. Sharov (1982). Systems and Models. Radio and Communications, Moscow, 1982, 152 pp.

ON A SIMILARITY RATIO*

Elena Castiñeira†
Dept. Matemática Aplicada
Universidad Politécnica de Madrid
28660 Boadilla del Monte
ecastineira@fi.upm.es

Susana Cubillo
Dept. Matemática Aplicada
Universidad Politécnica de Madrid
28660 Boadilla del Monte
scubillo@fi.upm.es

Enric Trillas
Dept. Inteligencia Artificial
Universidad Politécnica de Madrid
28660 Boadilla del Monte
etrillas@fi.upm.es

Summary

In [3] an index of similarity among concepts, playing an important role in the Analogical Case-Based Reasoning there shown, is introduced. That index is here studied as a fuzzy relation by analysing for which continuous t-norms T is a T-indistinguishability. That index is also considered from a metric viewpoint.

Furthermore, a series of dues are given with regard to the possible choice of a numerical threshold α best suited for considering two concepts A and B as similar when their index of similarity exceeds the value α.

Key words: Similarity ratio, fuzzy relation, T-indistinguishability.

1. The mathematical study of fuzzy equivalences, or T-indistinguishabilities, has led to the publication of a good many papers (see [9], [14], [5], etc.). It is worthy of note that, within Theoretical Fuzzy Logic, the mathematical analysis of T-indistinguishabilities is a field that is important in itself and of undeniable interest for fuzzy logic applications, such as control theory, for instance (see [6], etc.). Furthermore, there are papers in which given indices are used to measure similarity and where fuzzy logic concepts, that could possibly reinforce the theoretical framework of the proposed methodology, are not taken into account. This is the case of [3], an interesting paper in which what is referred to as the "ratio of similarity" between two concepts A and B, with respective numerical signatures $(a_1,\ldots,a_n)$ and $(b_1,\ldots,b_n)$ (that can be easily considered as belonging to $[0,1]^n$) related to an n-tuple $P=\{p_1,\ldots,p_n\}$ of relevant properties of the domain in question, plays a decisive role. In the above paper, the **ratio of similarity** is taken to be the number

$$\mathrm{sim}\,(A,B)=\frac{\sum_{i=1}^{n}\mathrm{Min}(a_i,b_i)}{\mathrm{Max}\left(\sum_{i=1}^{n}a_i,\sum_{i=1}^{n}b_i\right)}\in[0,1], \quad (1)$$

which defines the crisp relations of analogy given by $A\,\mathbf{sim}_\alpha\,B$ if and only if $\mathrm{sim}(A,B)\geq\alpha$. The function "sim" can evidently be viewed as a fuzzy relation.

2. Let $E=\{x_1,\ldots,x_n\}$ be a finite universe of discourse and $\mathcal{F}(E)=[0,1]^E$ be the set of all the fuzzy subsets of E. The fuzzy relation $\mathbf{sim}:\mathcal{F}(E)\times\mathcal{F}(E)\to[0,1]$, defined by

$$\mathbf{sim}(A,B)=\frac{\sum_{i=1}^{n}\mathrm{Min}(A(x_i),B(x_i))}{\mathrm{Max}\left(\sum_{i=1}^{n}A(x_i),\sum_{i=1}^{n}B(x_i)\right)},$$

is the above-mentioned formula (1), bearing in mind that $A(x_i)=a_i$ and $B(x_i)=b_i$. Obviously, $\mathbf{sim}(A,B)\in[0,1]$ for any A and B of $\mathcal{F}(E)$. Also, $\mathbf{sim}(A,A)=1$ and $\mathbf{sim}(A,B)=\mathbf{sim}(B,A)$ for any A and B of $\mathcal{F}(E)$; that is, **sim** is a reflexive and symmetrical fuzzy relation in $\mathcal{F}(E)$ (see [7]). Furthermore, it can be proved that it is a separating relation, that is,

Theorem 1 $\mathbf{sim}(A,B)=1$ if and only if $A=B$.

Note also that $\mathbf{sim}(A,B)=0$ if and only if $A(x_i)=0$ or $B(x_i)=0$ is satisfied for every $i\in\{1,\ldots,n\}$. That is to say

Theorem 2 It is $\mathbf{sim}(A,B)=0$ if and only if $A\cap B=\varphi_\emptyset$, by defining as usual $(A\cap B)(x_i)=\mathrm{Min}(A(x_i),B(x_i))$ for every $x_i\in E$.

Were **sim** to be a T-indistinguishability for a t-norm T **sim** would have to be T-transitive (see [14]), that is, $T(\mathbf{sim}(A,B),\mathbf{sim}(B,C))\leq\mathbf{sim}(A,C)$ for any A, B and C of $\mathcal{F}(E)$. Let $\mathcal{F}(\mathrm{Prod})$ be the family of t-norms of the product, that is, the t-norms $T_\varphi=\varphi^{-1}\circ\mathrm{Prod}\circ(\varphi\times\varphi)$, where $\varphi:[0,1]\to[0,1]$ is continuous, strictly increasing and such that $\varphi(0)=0$ and $\varphi(1)=1$. Let us see , firstly, that if $T\in\mathcal{F}(\mathrm{Prod})$ ó $T=\mathrm{Min}$, then **sim** is not T-transitive. To this end,

Lemma 3 If **sim** is T-transitive, then $T(a,b)=0$ for every $(a,b)\in[0,1]^2$ such that $a+b\leq1$.

*Research funded by the Spanish Interministerial Commission of Science and Technology (CICYT), projects n° TIC96-1393-C06 and n° TIC98-0996-C02-02

†Corresponding author

Theorem 4 If $T = \text{Min}$ or $T \in \mathcal{F}(\text{Prod})$, **sim** is not T-indistinguishability.

Proof. Suffice it to take a and b such that $0 < a, b \leq 1/2$, then $a+b \leq 1$ and, nevertheless, $\text{Min}(a,b) > 0$ and $T(a,b) = \varphi^{-1}(\varphi(a) \cdot \varphi(b)) > \varphi^{-1}(0) = 0$. ■ However,

Theorem 5 **sim** is a W-indistinguishability.

Proof: Let A, B and C of $\mathcal{F}(E)$, where $E = \{x_1, \ldots, x_n\}$ and let us denote $a_i = A(x_i)$, $b_i = B(x_i)$ and $c_i = C(x_i)$, for each $i \in \{1, \ldots, n\}$, and $a = \sum_{i=1}^n a_i$, $b = \sum_{i=1}^n b_i$ and $c = \sum_{1=1}^n c_i$. It is a matter of proving that $W(\mathbf{sim}(A,B), \mathbf{sim}(B,C)) \leq \mathbf{sim}(A,C)$ or, alternatively, that

$$\mathbf{sim}(A,B) + \mathbf{sim}(B,C) \leq \mathbf{sim}(A,C) + 1 \qquad (2)$$

Suppose, firstly, that $b \geq \text{Max}(a,c)$; in this case, the inequality to be proven is:

$$\frac{\sum_{i=1}^n \text{Min}(a_i,b_i)}{b} + \frac{\sum_{i=1}^n \text{Min}(b_i,c_i)}{b} \leq \frac{\sum_{i=1}^n \text{Min}(a_i,c_i)}{\text{Max}(a,c)} + 1.$$

As $\text{Min}(a_i,b_i) + \text{Min}(b_i,c_i) \leq a_i + b_i$ and $\text{Min}(a_i,b_i) + \text{Min}(b_i,c_i) \leq b_i + c_i$ for every $i \in \{1,\ldots,n\}$, then $\text{Min}(a_i,b_i) + \text{Min}(b_i,c_i) \leq \text{Min}(a_i,c_i) + b_i$, hence, by summing in i and dividing by b, we get

$$\frac{\sum_{i=1}^n \text{Min}(a_i,b_i)}{b} + \frac{\sum_{i=1}^n \text{Min}(b_i,c_i)}{b} \leq \frac{\sum_{i=1}^n \text{Min}(a_i,c_i)}{b} + 1,$$

and as $\frac{\sum_{i=1}^n \text{Min}(a_i,c_i)}{b} \leq \frac{\sum_{i=1}^n \text{Min}(a_i,c_i)}{\text{Max}(a,c)}$ we get the inequality we sought to prove.

Suppose, secondly, that $a \geq \text{Max}(b,c)$. Consider the following partition of the set of indices:

$$\begin{aligned} I &= \{i \in \{1,\ldots,n\}; a_i \geq b_i \geq c_i\} \\ J &= \{i \in \{1,\ldots,n\}; a_i > c_i > b_i\} \\ K &= \{i \in \{1,\ldots,n\}; b_i > a_i > c_i\} \\ L &= \{i \in \{1,\ldots,n\}; b_i > c_i \geq a_i\} \\ M &= \{i \in \{1,\ldots,n\}; c_i \geq a_i > b_i\} \\ N &= \{i \in \{1,\ldots,n\}; c_i > b_i \geq a_i\} \end{aligned}$$

and say $a_I = \sum_{i \in I} a_i$, $b_I = \sum_{i \in I} b_i$, $c_I = \sum_{i \in I} c_i$ and so on for each element of the partition. Using this notation, we get the relations

$$\begin{array}{lll} a_I \geq b_I \geq c_I & b_K > a_K > c_K & c_M \geq a_M > b_M \\ a_J > c_J > b_J & b_L > c_L \geq a_L & c_N > b_N \geq a_N \end{array} \qquad (3)$$

and then $\mathbf{sim}(A,B) + \mathbf{sim}(B,C) = \frac{b_I+b_J+a_K+a_L+b_M+a_N}{a} + \frac{c_I+b_J+c_K+c_L+b_M+b_N}{\text{Max}(b,c)} \leq \frac{b_I+b_J+a_K+a_L+b_M+a_N}{a} + \frac{c_I+b_J+c_K+c_L+b_M+b_N}{b}$.

If the last member of the above expression is found to be less than or equal to

$$\mathbf{sim}(A,C) + 1 = \frac{c_I + c_J + c_K + a_L + a_M + a_N}{a} + 1$$

we will have proven that the inequality (2) holds in the event that $a \geq \text{Max}(b,c)$. However, this is tantamount to proving

$$\frac{b_I + b_J + a_K + b_M}{a} + \frac{c_I + b_J + c_K + c_L + b_M + b_N}{b} \leq \frac{c_I + c_J + c_K + a_M}{a} + 1 \text{ that amounts to}$$

$$\begin{aligned} &b_I b + b_J b + a_K b + b_M b + c_I a + c_K a + c_L a \leq \\ &c_I b + c_J b + c_K b + a_M b + b_I a + b_K a + b_L a. \end{aligned} \qquad (4)$$

From (3) it is

$$b_J b + b_M b + c_L a \leq c_J b + a_M b + b_L a; \qquad (5)$$

also, since $b \leq a$ and $b_I - c_I \geq 0$, it is

$$b_I b + c_I a \leq b_I a + c_I b; \qquad (6)$$

and since $c_K < a_K < b_K$ it is

$$a_K b + c_K a \leq b_K a + c_K b. \qquad (7)$$

By summing (5), (6) and (7) member by member, we get the inequality (4).

Finally, suppose that $c \geq \text{Max}(a,b)$. Then, as in the above case, $\mathbf{sim}(C,B) + \mathbf{sim}(B,A) \leq \mathbf{sim}(C,A) + 1$ and since **sim** is symmetric, (2) again holds; hence, **sim** is W-transitive.■

3. The fuzzy relation **sim** is not only T-transitive for the Lukasiewicz t-norm $T = W$, there also exist t-norms T_φ of the Lukasiewicz family, $\mathcal{F}(W)$ (that is, $T_\varphi = \varphi^{-1} \circ W \circ (\varphi \times \varphi)$ and φ as above when the family of the Product, $\mathcal{F}(\text{Prod})$, was introduced before Lemma 3) for which **sim** is T_φ-transitive. In fact, it is easy to prove

Theorem 6 If $T_\varphi \in \mathcal{F}(W)$ is such that $T_\varphi(x,y) \leq W(x,y)$ for each $(x,y) \in [0,1]^2$, **sim** is a T_φ-indistinguishability.

For example, if $\varphi_n(x) = x^n$, then $T_{\varphi_n} = \varphi_n^{-1} \circ W \circ (\varphi_n \times \varphi_n) \leq W$ in $[0,1]^2$ and then **sim** is T_{φ_n}-transitive and a T_{φ_n}-indistinguishability for any natural number n.

Theorem 7 If $T_\varphi \in \mathcal{F}(W)$ and **sim** is a T_φ-indistinguishability, then $N_{id}(x) \leq N_\varphi(x)$ for every $x \in [0,1]$, where N_{id} and N_φ are the negations associated with the identity and with φ, respectively.

Proof. If an $a \in [0,1]$ existed such that $N_\varphi(a) = \varphi^{-1}(1-\varphi(a)) < N_{id}(a) = 1-a$, then we could consider b such that $N_\varphi(a) < b < 1-a$, hence $\varphi(a) + \varphi(b) >$

$\varphi(a)+\varphi(N_\varphi(a)) = \varphi(a)+\varphi(\varphi^{-1}(1-\varphi(a))) = 1$; and, therefore, $T_\varphi(a,b) = \varphi^{-1}(\varphi(a)+\varphi(b)-1) > \varphi^{-1}(0) = 0$ which contradicts lemma 3 since $a+b<1$. ■ For the characterization of Strong-Negations, see [10].

As an application of the above result, consider that if $\varphi(x) = x^\alpha$, where $\alpha < 1$ is a real number, then **sim** is not T_φ-transitive. Since $x^\alpha \geq x$ if and only if $(1-x)^{1/\alpha} \geq (1-x^\alpha)^{1/\alpha} = N_\varphi(x)$, it is $N_\varphi(x) \leq (1-x)^{1/\alpha} < 1-x = N_{id}(x)$ if $x \in (0,1)$. Therefore, **sim** cannot be T_φ-transitive.

Note: It is easy to prove that for any function φ that satisfies $\varphi(x) \leq x$ for every $x \in [0,1]$, it is $N_{id} \leq N_\varphi$

The reciprocal of last theorem does not hold. Indeed, the function $\varphi : [0,1] \to [0,1]$ defined by

$$\varphi(x) = \begin{cases} x & \text{if } x \geq 4/5 \\ \frac{5}{4}x^2 & \text{if } 0 \leq x < 4/5 \end{cases}$$

is continuous, strictly increasing and satisfies $\varphi(0) = 0$, $\varphi(1) = 1$; additionally $\varphi(x) \leq x$ holds for every $x \in [0,1]$. Then, taking into account the above observation, is $N_{id} \leq N_\varphi$. Let us see, however, that **sim** is not T_φ-transitive.

Let $E = \{x_1, x_2, x_3\}$ and consider A, B and C, defined as shown by the table below:

A	B	C
1/5	2/5	2/5
2/5	2/5	2/5
2/5	1/5	0
$\sum A(x_i) = 1$	$\sum B(x_i) = 1$	$\sum C(x_i) = 4/5$

Then, as $\mathbf{sim}(A,B) = 4/5$, $\mathbf{sim}(B,C) = 4/5$ and $\mathbf{sim}(A,C) = 3/5$ it is $T_\varphi(\mathbf{sim}(A,B), \mathbf{sim}(B,C)) = \varphi^{-1}(\varphi(4/5)+\varphi(4/5)-1) = \varphi^{-1}(3/5) > \mathbf{sim}(A,B)$.

4. Like any fuzzy set, the relation **sim** has its associated α-cuts (see. [7]); that is, the classical relations $\mathbf{sim}_\alpha$ in $\mathcal{F}(E)$, defined by

$$A\ \mathbf{sim}_\alpha\ B \text{ if and only if } \mathbf{sim}(A,B) \geq \alpha.$$

Such relations are referred to in [3] as relations of **analogy** and, obviously, are all reflexive and symmetrical, although the only one that is transitive and, therefore, a relation of equivalence, corresponds to $\alpha = 1$.

Theorem 8 $\mathbf{sim}_\alpha$ is transitive if and only if $\alpha = 1$.

Proof. If $\alpha = 1$, it follows from theorem 1 that $\mathbf{sim}_\alpha$ is the relation of equality and is, therefore, transitive. If $\mathbf{sim}_\alpha$ is transitive and suppose that $\alpha < 1$, consider ε, $0 < \varepsilon < \alpha$, and A, B, $C \in \mathcal{F}(E)$, where $E = \{x_1, x_2, x_3\}$, defined according by the table:

A	B	C
α	α	$\alpha-\varepsilon$
0	$1-(\alpha+\varepsilon)$	0
0	ε	$1-(\alpha-\varepsilon)$
$\sum A(x_i) = \alpha$	$\sum B(x_i) = 1$	$\sum C(x_i) = 1$

Hence $\mathbf{sim}(A,B) = \alpha$, $\mathbf{sim}(B,C) = \alpha$, as $\varepsilon \leq 1-(\alpha-\varepsilon)$ if and only if $\alpha \leq 1$, and $\mathbf{sim}(A,C) = \alpha-\varepsilon < \alpha$; then, $\mathbf{sim}_\alpha$ is not transitive. ■ It is clear, pursuant to theorem 1 that $\mathbf{sim}_1 = \{(A,A); A \in \mathcal{F}(E)\}$, whereas $\mathbf{sim}_0 = \mathcal{F}(E) \times \mathcal{F}(E)$ and, pursuant to theorem 2 it is $\mathbf{sim}_0 - \left(\bigcup_{\alpha>0} \mathbf{sim}_\alpha\right) = \{(A,B) \in \mathcal{F}(E) \times \mathcal{F}(E); A \cap B = \varphi_\emptyset\}$.

5. The fact that the fuzzy relation **sim** is T-transitive for some t-norms $T = \varphi^{-1} \circ W \circ (\varphi \times \varphi)$ means that the question can be addressed from a metric viewpoint, thanks to the following result

Lemma 9 The necessary and sufficient condition for a fuzzy relation $S : E \times E \to [0,1]$ to be a T-indistinguishability for $T \in \mathcal{F}(W)$ where $T = \varphi^{-1} \circ W \circ (\varphi \times \varphi)$, is that the function $D : E \times E \to [0,1]$ defined by $D = 1 - \varphi \circ S$, is a distance.

The case $\varphi = \mathrm{id}$ was considered in [14] and the general case in [12]. Consequently,

Theorem 10 If $T \in \mathcal{F}(W)$, with $T = \varphi^{-1} \circ W \circ (\varphi \times \varphi)$, the necessary and sufficient condition for the fuzzy relation **sim** to be a T-indistinguishability is that the function $D_\varphi = 1 - \varphi \circ \mathbf{sim}$ is a distance.

Moreover, D_φ is "separating", in the sense that $D_\varphi(A,B) = 0$ if and only if $A = B$, follows from theorem 1.

If $T = \varphi^{-1} \circ W \circ (\varphi \times \varphi)$ is a t-norm of $\mathcal{F}(W)$ for which **sim** is T-transitive, it is obvious that

$$A\ \mathbf{sim}_\alpha\ B \quad \text{if and only if} \quad D_\varphi(A,B) \leq 1-\varphi(\alpha).$$

holds, and the neighbourhoods

$$E_{1-\varphi(\alpha)}(A) = \{B \in \mathcal{F}(E); D_\varphi(A,B) \leq 1-\varphi(\alpha)\},$$

in the respective metric space $(\mathcal{F}(E), D_\varphi)$ afford the equivalent definitions:

$$A\ \mathbf{sim}_\alpha\ B \ \textbf{iff}\ \ B \in E_{1-\varphi(\alpha)}(A) \ \textbf{iff}\ \ A \in E_{1-\varphi(\alpha)}(B).$$

That is, two functions A and B are linked by the classical relation $\mathbf{sim}_\alpha$ (are "analogical" with the threshold α, as stated in [3]) if and only if the distance between them is at most $1-\varphi(\alpha)$. Obviously, it is $E_{1-\varphi(\alpha)}(A) = \{A\}$ if and only if $\alpha = 1$ and, of course, the neighbourhoods $E_{1-\varphi(\alpha)}(A)$ are **clusters** of analogical functions.

Pursuant to the above, if $0 \leq \alpha \leq \beta \leq 1$, then $\mathbf{sim}_1 \subset \mathbf{sim}_\beta \subset \mathbf{sim}_\alpha \subset \mathbf{sim}_0$. That is, if the values

of the threshold α in $[0,1]$ are high, each $A \in \mathcal{F}(E)$ is analogical to only a few $B \in \mathcal{F}(E)$, whereas if they are low, it is similar to a lot. Therefore, it is reasonable to think that, as a general rule, the **best** relations of analogy $\mathbf{sim}_\alpha$ will correspond to intermediate values of the threshold α. Furthermore, as we are likely to want to evade logically incompatible pairs (A,B) (that is, such that $A \cap B = \varphi_\emptyset$, see [13]), the value for the threshold should be taken from the intermediate region closest to the value $\alpha = 1$; indeed, the paper referenced as [3] considers the threshold $\alpha = 0.7$ which the authors claim to have attained experimentally.

Besides classical relations $\mathbf{sim}_\alpha$, for $\alpha < 1$ are not transitive, they have a kind of restricted transitivity that can allow some control of the analogical reasoning systems in which they are used. In fact, as it is always

$$\mathbf{sim}(A,C) \geq W(\mathbf{sim}(A,B), \mathbf{sim}(B,C)),$$

inequalities $\mathbf{sim}(A,B) \geq \alpha$ and $\mathbf{sim}(B,C) \geq \alpha$ imply $\mathbf{sim}(A,C) \geq W(\alpha,\alpha)$, that is, a lower bound $W(\alpha,\alpha)$ of $\mathbf{sim}(A,C)$ is obtained. Nevertheless, if it is $W(\alpha,\alpha) = 0$ this inequality is not at all informative on the values of $\mathbf{sim}(A,C)$.

To have $W(\alpha,\alpha) = \mathrm{Max}(0, 2\alpha - 1) > 0$ it suffices to have $\alpha > 0.5$. Consequently, if $\alpha > 0.5$ the scheme:

$$\frac{A \ \mathbf{sim}_\alpha \ B \qquad B \ \mathbf{sim}_\alpha \ C}{A \ \mathbf{sim}_{W(\alpha,\alpha)} \ C,}$$

holds with $W(\alpha,\alpha) > 0$, and the classical relation $\mathbf{sim}_\alpha$ shows this restricted transitivity.

As it was proven beforehand, if $T_\varphi \in \mathcal{F}(W)$ verifies $T_\varphi \leq W$ the fuzzy relation **sim** is T_φ transitive, and then: $\mathbf{sim}(A,C) \geq W(\alpha,\alpha) \geq T_\varphi(\alpha,\alpha)$. When it is also $T_\varphi(\alpha,\alpha) > 0$, a small and then not too good lower bound than $W(\alpha,\alpha)$ is reached for $\mathbf{sim}(A,C)$ if $\alpha > 0.5$. Because of with $\varphi(x) = x^2$ it is $T_\varphi \in \mathcal{F}(W)$ and $T_\varphi \leq W$, the number $\sqrt{\mathrm{Max}(0, 2\alpha^2 - 1)}$ is such a bound provided it is positive, a situation that happens if $\alpha > \sqrt{0.5} \simeq 0.707107$. Then, it is enough to take $\alpha > 0.7072$ to be sure that the relation $\mathbf{sim}_\alpha$ enjoys the restricted transitivity shown by:

$$\frac{A \ \mathbf{sim}_\alpha \ B \qquad B \ \mathbf{sim}_\alpha \ C}{A \ \mathbf{sim}_{T_\varphi(\alpha,\alpha)} \ C,}$$

with $T_\varphi(\alpha,\alpha) = \sqrt{\mathrm{Max}(0, 2\alpha^2 - 1)} > 0$. Perhaps a refinement of the experiments that in [3] conducted to use $\alpha > 0.7$ can show that the ratio **sim** was taken there as T_φ-transitive with $\varphi(x) = x^2$ and not as W-transitive. In that case, A and B can be taken as "analogous" if the "distance" between both is less or equal than $1 - \varphi(\alpha) = 1 - (\sqrt{5})^2 = 0.5$

Each application is likely to call for an experimental determination of the maximum threshold. However, **an upper bound β of the maximum threshold can be determined** for each family $\{A_1, \ldots, A_n\}$ of functions as follows. Calculate $\mathbf{sim}(A_i, A_j) = \beta_{ij}$ for each pair of indices i, j, and let $\beta = \mathrm{Min}\{\beta_{ij}; 1 \leq i, j \leq n\}$; as $A_i \, \mathbf{sim}_\beta \, A_j$ for each pair A_i, A_j, then it is also $A_i \, \mathbf{sim}_\alpha \, A_j$ for any $\alpha \leq \beta$.

References

[1] Bouchon-Meunier, B. and Valverde, L., (1997)"A Resemblance Approach to Analogical Reasoning Functions" in *Fuzzy Logic for Artificial Intelligence: Towards Intelligent Systems*(T.P. Martin and A.L. Ralescu, eds.), pp. 266-272 (Springer-Verlag, Berlin)

[2] Chouraqui, E. and Inghilterra, C.(1996), "A model of Case-Based Reasoning for Solving Problems of Geometry in a Tutoring System ", in *Intelligent Learning Enviroments: The Case of Geometry* (edited by J-M Laborde),pp.1-16, Springer-Verlag (Berlin-Heidelberg).

[3] Jacas, J. and Recasens, J.(1995) "Fuzzy T-transitive relations: eigenvectors and generators". *Fuzzy Sets & Systems*,72, 147-154.

[4] Klawonn, F. and Castro, J.L.(1995) "Similarity in Fuzzy Reasoning", in *Mathware & Soft Computing*, 2, 197-228.

[5] Nguyen, H.T. and Walker, E.A.(1997), *A First Course in Fuzzy Logic*, CRC Press, Boca Raton.

[6] Trillas, E.(1993) "On Logic and Fuzzy Logic", in *Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*,1-2,107-137.

[7] Trillas, E. (1979) "Sobre funciones de negación en la teoría de conjuntos difusos". *Stochastica*, Vol.III, 1, 47-60.

[8] Trillas E., Castiñeira, E. and Pradera A. "On the Equivalence between Distances and T-Indistinguishabilities" (in these Proccedings).

[9] Trillas, E. and Cubillo, S. "On Non-Contradictory Input/Output Couples in Zadeh's CRI" (To appear in Proceedings NAFIPS'99).

[10] Trillas, E. y Valverde, L.(1984) "An Inquiry into Indistinguishability Operators". *Aspects of Vagueness* (Eds. H.J. Skala et altri), 231-256, Reidel Pubs.

[11] Zadeh, L.A.(1971) "Similarity relations and fuzzy orderings", in *Inform. Sci.* 3,177-200.

On Implicative Closure Operators in Approximate Reasoning

R. O. Rodríguez[1], F. Esteva[2], P. Garcia[2] and L. Godo[2]

[1] Dpto. de Computación, Fac. Ciencias Exactas y Naturales
Universidad de Buenos Aires, 1428 Buenos Aires, Argentina
ricardo@dc.uba.ar

[2] Institut d'Investigació en Intel·ligència Artificial, CSIC
Campus Univ. Autònoma Barcelona, 08193 Bellaterra, Spain
{esteva,pere,godo}@iiia.csic.es

Summary

This paper introduces a new definition of fuzzy closure operator called Implicative Closure Operators. The Implicative Closure Operators generalize some notions of fuzzy closure operators given by differents authors. We show that the Implicative Closure Operators capture some usual Consequence Relations used in Approximate Reasoning, like the Approximation and Proximity entailments defined by Dubois et al. [5] and the Natural Inference Operator defined by Boixader and Jacas [1].

Keywords: Implication Measures, Approximate Reasoning, Fuzzy Closure Operators, Closure Systems and Fuzzy Consequence Relations.

1 Introduction

In the last years, many works have been devoted to extend the notions of closure operators, closure systems and consequence relations from two valued logic to many valued logic. The best well-known approach to many-valued closure operators is due to Pavelka [11]. He defines such operators (in the standard sense of Tarski) as mappings from fuzzy sets of formulas to fuzzy sets of formulas, i.e.,

Definition 1 (Fuzzy Closure Operator)
A Fuzzy Closure Operator on a language $\mathcal{L}$ is a mapping $\tilde{C} : \mathcal{F}(\mathcal{L}) \mapsto \mathcal{F}(\mathcal{L})$ fulfilling, for all $A, B \in \mathcal{F}(\mathcal{L})$, the following properties:

$\tilde{C}1$) *fuzzy inclusion:* $A \subseteq \tilde{C}(A)$

$\tilde{C}2$) *fuzzy monotony: if* $A \subseteq B$ *then* $\tilde{C}(A) \subseteq \tilde{C}(B)$

$\tilde{C}3$) *fuzzy idempotence:* $\tilde{C}(\tilde{C}(A)) \subseteq \tilde{C}(A)$

where $(L, \wedge, \vee, 0, 1)$ is a complete lattice with minimum 0 and maximum 1 and $\mathcal{F}(\mathcal{L})$ denotes the the fuzzy power set (i.e. the set $L^{\mathcal{L}}$ of all fuzzy subsets) over a language $\mathcal{L}$.

On the other hand, Chakraborty extends in [4] the concept of consequence relation by defining Graded Consequence Relations as fuzzy relations between crisp sets of formulas and formulas.

Definition 2 (Graded Consequence Relation)
A fuzzy relation $g_c : \mathcal{P}(\mathcal{L}) \times \mathcal{L} \longmapsto L$ is called a Graded Consequence Relation if, for every $A, B \in \mathcal{P}(\mathcal{L})$ and $p, q \in \mathcal{L}$, g_c fulfills:

g_c1) *fuzzy reflexivity:* $g_c(A, p) = 1$ *for all* $p \in A$

g_c2) *fuzzy monotony: If* $B \subseteq A$ *then* $g_c(B, p) \leq g_c(A, p)$

g_c3) *fuzzy cut:* $(\inf_{q \in B} g_c(A, q)) \otimes g_c(A \cup B, p) \leq g_c(A, p)$, *where* $\otimes$ *is a t-norm operation on* L

where $\mathcal{P}(\mathcal{L})$ denotes the classical power set over a language $\mathcal{L}$

In [8] Gerla examines the links between Fuzzy Closure Operators and Graded Consequence Relations. Castro et al. point out in [3] that several methods of approximate reasoning used in Artificial Intelligence are not covered by Graded Consequence Relations and they introduce a more general concept of Fuzzy Consequence Relation.

Definition 3 (Fuzzy Consequence Relation)
Any fuzzy relation $\tilde{g} : \mathcal{F}(\mathcal{L}) \times \mathcal{L} \longmapsto L$ is called a Fuzzy Consequence Relation if the following three properties hold for every $A, B \in \mathcal{F}(\mathcal{L})$ and $p, q \in \mathcal{L}$:

$\tilde{g}1$) *fuzzy reflexivity:* $A(p) \leq \tilde{g}(A, p)$

$\tilde{g}2$) *fuzzy monotony: If* $B \subseteq A$ *then* $\tilde{g}(B, p) \leq \tilde{g}(A, p)$

$\tilde{g}3$) *fuzzy cut: If for all p,* $B(p) \leq \tilde{g}(A, p)$, *then for all q,* $\tilde{g}(A \cup B, q) \leq \tilde{g}(A, q)$

Notice that the Fuzzy Closure Operators and the Fuzzy Consequence Relations are related by the following properties:

- If $\tilde{C}$ is a Fuzzy Closure Operator then, $\tilde{g}(A,p) = \tilde{C}(A)(p)$ is a Fuzzy Consequence Relation.
- If $\tilde{g}$ is a Fuzzy Consequence Relation then, $\tilde{C}(A) = \tilde{g}(A,.)$ is a Fuzzy Closure Operator.

On the other hand a Fuzzy Closure Operator $\tilde{C}$ on $\mathcal{F}(\mathcal{L})$ induces a Closure System $\mathcal{C}$ on $\mathcal{F}(\mathcal{L})$ defined by

$$\mathcal{C} = \{T \in \mathcal{F}(\mathcal{L}) \mid \tilde{C}(T) = T\}.$$

And conversely, every closure system $\mathcal{C}$ in $\mathcal{F}(\mathcal{L})$ induces a fuzzy closure operator $\tilde{C}$ on $\mathcal{F}(\mathcal{L})$ defined by

$$\tilde{C}(A) = \bigwedge\{T \in \mathcal{C} \mid A \leq T\}.$$

In this paper we generalize, for fuzzy sets of formulas, Chakraborty's Graded Consequence Relations by defining what we call Implicative Consequence Relations, which generalize, at the same time, Castro et al.'s Fuzzy Consequence Relations. From these Implicative Consequence Relations we introduce and study their corresponding Implicative Closure Operators and Closure Systems. Finally we show that Implicative Consequence Relations also capture some well-known approximate entailments like the Approximation and Proximity entailments introduced by Dubois et al. in [5] and the Natural Inference Operator defined by Boixader and Jacas in [1].

2 Implicative Closure Operators

We are interested in a generalization of Graded Consequence Relations over fuzzy sets of formulas. Therefore, we introduce a new kind of fuzzy consequence relations that will be called *Implicative Consequence Relations*, since we will give the Fuzzy Cut property by means of an implication operation $\rightarrow$ on L. Thus we extend the lattice $(L, \wedge, \vee, 0, 1)$ to a complete BL-algebra[1], i.e. a structure $(L, \wedge, \vee, \otimes, \rightarrow, 0, 1)$ where $(L, \wedge, \vee, 0, 1)$ is a complete distributive lattice, $(L, \otimes, 1)$ is a commutative monoid, $(\otimes, \rightarrow)$ is a residuated pair, i.e. they verify

$$x \otimes y \leq z \text{ if and only if } x \leq y \rightarrow z,$$

and for all $x, y, z \in L$, it holds that $x \wedge y = x \otimes (x \rightarrow y)$ and $(x \rightarrow y) \vee (y \rightarrow x) = 1$. The operation $\rightarrow$ is usually called *residuated implication*.

[1] BL-algebras are introduced by Hájek in [9] as the algebraic counterpart of the so-called Basic Fuzzy Logic, which is the logic of continuous t-norms.

Definition 4 (Implicative Cons. Relation) *A fuzzy relation $\tilde{g}_c : \mathcal{F}(\mathcal{L}) \times \mathcal{L} \longmapsto L$ is called an Implicative Consequence Relation if, for every $A, B \in \mathcal{F}(\mathcal{L})$ and $p, q \in \mathcal{L}$, $\tilde{g}_c$ fulfills:*

$\tilde{g}_c 1)$ *fuzzy reflexivity:* $A(p) \leq \tilde{g}_c(A,p)$

$\tilde{g}_c 2)$ *fuzzy monotony: If* $B \subseteq A$ *then* $\tilde{g}_c(B,p) \leq \tilde{g}_c(A,p)$

$\tilde{g}_c 3)$ *fuzzy cut:* $[\inf_{q \in \mathcal{L}}(B(q) \rightarrow \tilde{g}_c(A,q))] \otimes \tilde{g}_c(A \cup B, p) \leq \tilde{g}_c(A,p)$ [2]

It is easy to prove that Graded Consequence Relations are a particular case of Implicative Consequence Relations. In fact, the main difference is to consider a graded fuzzy inclusion instead of an strict fuzzy inclusion in the cut property.

Implicative Consequence Relations also have a representation theorem which is a generalization of the one given in [4] for Graded Consequence Relations.

Theorem 1 *A fuzzy relation $\tilde{g}_c : \mathcal{F}(\mathcal{L}) \times \mathcal{L} \longmapsto L$ is an Implicative Consequence Relation if, and only if, there exists a family of fuzzy sets $\{T_i\}_{i \in I}$ satisfying $\tilde{g}_c(A,p) = \inf_i\{[\inf_{q \in \mathcal{L}} A(q) \rightarrow T_i(q)] \rightarrow T_i(p)\}$.*

The fuzzy sets $\{T_i\}_{i \in I}$ defining the Implicative Consequence Relation $\tilde{g}_c$, in the sense of the above theorem, will be called *generators* of $\tilde{g}_c$.

From the above theorem, it is easy to prove that any Implicative Consequence Relation satisfies the next additional property:

$\tilde{g}4)$ $\tilde{g}_c(A \otimes \bar{k}, p) \geq \tilde{g}_c(A,p) \otimes k$,

where $\bar{k}$ is the constant fuzzy set such that all the elements of $\mathcal{L}$ belong to it with value k.

The Fuzzy Closure Operators corresponding to Implicative Consequence Relations, which we shall call them *Implicative Closure Operators*, are characterized by the properties: $\tilde{C}1, \tilde{C}2, \tilde{C}3$, and

$\tilde{C}4)$ $\tilde{C}_c(A \otimes \bar{k})(p) \geq \tilde{C}_c(A)(p) \otimes \bar{k}$.

Property $\tilde{C}4)$ is already mentioned by Gerla in [8] for the case when A is a crisp set and $\otimes = \min$. The following theorem characterizes *Implicative Closure System* corresponding to Implicative Closure Operators.

Theorem 2 *A closure system $\mathcal{C}_{\rightarrow}$ in $\mathcal{F}(\mathcal{L})$ is an Implicative Closure System if and only if, for any $E \in \mathcal{C}_{\rightarrow}$ and for any $k \in L$, $\bar{k} \rightarrow E \in \mathcal{C}_{\rightarrow}$.*

[2] This axiom could be also presented as $\tilde{g}_c(A \cup B, p) \leq (\inf_{q \in \mathcal{L}}(B(q) \rightarrow \tilde{g}_c(A,q))) \rightarrow \tilde{g}_c(A,p)$

We end this section by mentioning that Implicative Closure Operators satisfy a further property:

$\tilde{C}5$) *Coherence*: $\tilde{C}_c(A)(q) \geq \tilde{C}_c(\{p\})(q) \otimes A(p)$,

where $\{p\}$ denotes a singleton, i.e. $\{p\}(p) = 1$ and $\{p\}(q) = 0$ otherwise. This property already appears in [2] where it is proved that if a closure operator satisfies $\tilde{C}5$ then $\tilde{C}_c(\{p\})(q)$ defines a fuzzy preorder in $\mathcal{L}$. Actually, it is easy to prove the following proposition.

Proposition 1 *If $\tilde{C}_c$ is an Implicative Closure Operator with generators $\{T_i\}_{i\in I}$, then it holds*

$$\tilde{C}_c(\{p\})(q) = \inf_{i\in I}\{T_i(p) \to T_i(q)\}.$$

Notice that the family $\{T_i\}_{i\in I}$ is a family of generators of the fuzzy preorder $R(p \mid q) = \tilde{C}_c(\{p\})(q)$.

3 Closure Operators defined by a fuzzy preorder

Different authors (see for instance [2]) have studied the so-called Fuzzy Closure Operators defined by a fuzzy preorder.

Definition 5 *Given a fuzzy preorder $R : \mathcal{L} \times \mathcal{L} \longmapsto L$ on the language $\mathcal{L}$, the associated Fuzzy Closure Operator $\tilde{C}_R$ is defined by:*

$$\tilde{C}_R(A)(q) = \bigvee_{p\in\mathcal{L}}\{R(q \mid p) \otimes A(p)\}.$$

The mapping $\tilde{C}_R$ provides, for every fuzzy set h , the least extensional fuzzy set (w.r.t. R) containing h, and it is proved in [10] that it satisfies the following properties:

- **C1**: $h \leq \tilde{C}_R(h)$
- **C2**: $\tilde{C}_R(\bigvee_{i\in I} h_i) = \bigvee_{i\in I} \tilde{C}_R(h_i)$
- **C3**: $\tilde{C}_R \circ \tilde{C}_R = \tilde{C}_R$
- **C4**: $\tilde{C}_R(\bar{k}) = \bar{k}$
- **C5$^\otimes$**: $\tilde{C}_R(\{q\} \otimes \bar{k}) = \tilde{C}_R\{q\} \otimes \bar{k}$
- **C6$^\otimes$**: $\tilde{C}(h \otimes \bar{k}) = \tilde{C}(h) \otimes \bar{k}$

where $\{q\}$ denotes a singleton and $\bar{k}$ the constant fuzzy set $\bar{k}(q) = k$, for all $q \in \mathcal{L}$.

The following remarks can be easily proved:

- **C2** implies fuzzy inclusion property
- in the presence of **C2**, **C5$^\otimes$** is equivalent to **C6$^\otimes$**
- **C6$^\otimes$** implies **C4**

In [7] Esteva et al. prove that a Fuzzy Closure Operator is generated by a fuzzy preorder (in the sense of above definition) if, and only if, it satisfies **C1-C5$^\otimes$**. On the other hand the closure system associated to these Fuzzy Closure Operators is the set of generators of the fuzzy pre-order R (see [10]). This closure system is characterized by the following properties:

1) it is closed under arbitrary unions, and

2) for any fuzzy set $E \in \mathcal{C}_{\to}$ and for any element $k \in L$, it holds that $\bar{k} \to E$, $\bar{k} \otimes E \in \mathcal{C}_{\to}$.

Taking into account this characterization and the above definition, the following theorem holds.

Theorem 3 *A Fuzzy Closure Operator defined by a fuzzy preorder R is the minimal Implicative Closure Operator $\tilde{C}_c$ such that $\tilde{C}_c(\{q\})(p) = R(q \mid p)$, for any $p, q \in \mathcal{L}$.*

4 Implicative Closure Operators and aproximate reasoning

In [6], given a similarity relation S on the set Ω of boolean interpretations of a propositional language $\mathcal{L}$, a fuzzy set p^* on Ω, called *approximately* p, is associated to each proposition $p \in \mathcal{L}$ in the following way:

$$p^*(w) = \sup_{w'\models p} S(w, w').$$

From this definition, Dubois et al. define in [5] two graded consequence relations on $\mathcal{L} \times \mathcal{L}$.

Definition 6 (Approximate Cons. Relation) *For each $p, q \in \mathcal{L}$, we define $p \models^\alpha q$ iff $I_S(q \mid p) = \inf_{w\models p} \sup_{w'\models q} S(w, w') \geq \alpha$.*

Definition 7 (Proximity Cons. Relation) *For each $p, q \in \mathcal{L}$ and each $K \subseteq \mathcal{L}$, we define $p \models^\alpha_K q$ iff $J_{S,K}(q \mid p) = \inf_{w\models K} p^*(w) \to q^*(w) \geq \alpha$.*

In [5] it is proved that $I_S(q \mid p) = J_{S,\top}(q \mid p)$, where $\top$ stands for a boolean tautology, i.e. any formula whose set of models is the whole set Ω. In our framework, the Approximate and Proximity Consequence Relations can be obtained as Implicative Closure Operators. Indeed, consider the family of fuzzy sets $F = \{\tilde{w}\}_{w\in\Omega}$, where for each $w \in \Omega$, the fuzzy set $\tilde{w} : \mathcal{L} \longmapsto [0, 1]$ is defined by $\tilde{w}(q) = q^*(w)$. Now define, for all $A \in [0, 1]^{\mathcal{L}}$,

$$\tilde{C}^*_c(A)(q) = \inf_{\tilde{w}\in F}(\inf_{p\in\mathcal{L}}(A(p) \to \tilde{w}(p))) \to \tilde{w}(q)).$$

An easy computation shows that $\tilde{C}_c^*(\{p\})(q) = J_{S,\top}(q \mid p) = I_S(q \mid p)$, and we obtain the Approximate Consequence Relation. Moreover, if we consider now the family of fuzzy sets to be $F = \{\tilde{w}\}_{w \models K}$, for a subset $K \subseteq \mathcal{L}$, then what we obtain is $\tilde{C}_c^*(\{p\})(q) = J_{S,K}(q \mid p)$, that is, the Proximity Consequence Relation.

In [1] Boixader and Jacas analyze "...approximate reasoning through extensionality with respect to the natural $\otimes$-indistinguishability operator, by considering the indistinguishability level between fuzzy sets as a formal measure of their degree of similarity,...". For this purpose, they introduce a family of operators $I : [0,1]^U \longmapsto [0,1]^V$, where U and V are universes of discourse. These operators are called *extensional inference operators* if they preserve the pointwise order, i.e. if $A_1 \leq_U A_2$ then $I(A_1) \leq_V I(A_2)$, and satisfy $\otimes$-extensionality, i.e. $\bar{E}_U^{\otimes}(A_1, A_2) \leq \bar{E}_V^{\otimes}(I(A_1), I(A_2))$, where $\bar{E}_U^{\otimes}$ is the similarity function defined by

$$\bar{E}_U^{\otimes}(A_1, A_2) = \inf_{x \in U} A_1(x) \vee A_2(x) \to A_1(x) \wedge A_2(x)$$

and analogously for $\bar{E}_V^{\otimes}$. Moreover, they prove that it is possible to associate to any fuzzy rule *"If A then B"* the so-called *Natural Inference Operator*, which is the optimal from the extensionality point of view.

Definition 8 (Natural Inference Operator) *Given the rule "If A then B" with $A \in [0,1]^U$ and $B \in [0,1]^V$, the Natural Inference Operator $\bar{I}_{AB} : [0,1]^U \longmapsto [0,1]^V$ associated to this rule is defined by*

$$\bar{I}_{AB}(A')(v) = [\inf_{u \in U}(A'(u) \to A(u))] \to B(v).$$

Theorem 4 *([1, Theorem 15]) The Natural Inference Operator associated to the rule "If A then B" is the least specific inference operator satisfying:*

- *$\bar{I}_{AB}$ is extensional,*
- *$\bar{I}_{AB}(A) = B$,*
- *$\bar{I}_{AB}(A') \geq_V B$, for any $A' \in [0,1]^U$. Moreover, if $A' \leq_U A$, then $\bar{I}_{AB}(A') = B$*

Now we are in position to show that Implicative Closure Operators $\tilde{C}_c$ are Natural Inference Operators. In fact, for any $\tilde{C}_c$ there exists a family $\{T_i\}_{i \in I}$ such that

$$\tilde{C}_c(A)(p) = \inf_{i \in I}\{[\inf_{q \in \mathcal{L}} A(q) \to T_i(q)] \to T_i(p)\}$$

If we take $U = V = \mathcal{L}$, it is obvious that, for each T_i, the operator I_i defined as

$$I_i(A)(p) = [\inf_{q \in \mathcal{L}} A(q) \to T_i(q)] \to T_i(p)$$

is the Natural Inference Operator on $\mathcal{L}$ associated to the rule *"if T_i then T_i"*, and thus the following theorem holds.

Theorem 5 *Implicative Closure Operators are Natural Inference Operators.*

Moreover, our representation theorem 1 should be seen as a particular case of the representation theorem in [1]. Thus, the operators $\tilde{C}_c$ are always extensional in the above sense. Finally, remark that closure operators coming from a $\otimes$-transitive fuzzy preorders are obviously $\otimes$-extensional. But, in general, we can only assure that closure operators are min-extensional.

Proposition 2 *Let $\tilde{C}$ be the closure operator defined by the family of fuzzy sets $\{T_i\}_{i \in I}$ by $\tilde{C}(A)(p) = \inf\{T_i(p) \mid A \subseteq T_i\}$. Then $\tilde{C}$ is* min*-extensional.*

However, this result is not true in general for t-norms different of min, indeed there are examples that are not Lukasiewicz-extensional.

References

[1] D. Boixader, J. Jacas (1998) Extensionality based approximate reasoning. International Journal of Approximate Reasoning",19, pag. 221-230.

[2] J.L. Castro, E. Trillas (1999). Tarki's fuzzy consequence. Submitted to IJAR

[3] J.L. Castro , E. Trillas , S. Cubillo (1994). On consequence in approximate reasoning. Journal of Applied Non-Classical Logics, vol. 4 nro. 1, pp. 91-103.

[4] M.K. Chakraborty (1988). Use of fuzzy set theory in introducing graded consequence in multiple valued logic. Fuzzy Logic in Knowledge Systems, Decision and Control, M.M. Gupta and T.Yamakawa (eds). North Holland, Amsterdam, pp. 247-257.

[5] D. Dubois, F. Esteva, P. Garcia, L. Godo, H. Prade (1997). A Logical approach to interpolation based on similarity relations. Int. J. of Approximate Reasoning, 17 (1), 1-36.

[6] F. Esteva, P. Garcia, L. Godo (1994). Relating and extending semantical approaches to possibilistic reasoning. Int. J. of Approximate Reasoning, 11, 312-344.

[7] F. Esteva, P. Garcia, L. Godo, R.O. Rodríguez (1998). Fuzzy Approximation Relations, Modal Structures and Possibilistic Logic. Mathware & Softcomputing, 5, n. 2-3, 151-166.

[8] G. Gerla (1996). Graded Consequence Relations and Fuzzy Closure Operators. Journal of Applied Non-Clasical Logics, vol. 6 nro. 4, 369-379.

[9] P. Hájek, Methamatematics of Fuzzy Logic, Kluwer, 1998.

[10] J. Jacas, J. Recasens (1995). Fuzzy T-transitive relations: eigenvectors and generators. Fuzzy Sets & Systems 72, 147-154.

[11] J. Pavelka (1979) On the fuzzy logic I: Many valued rules and inference. Math. Logik Grundlag. Math. 25, 45-52.

ON (S,N)-IMPLICATIONS IN FUZZY LOGIC CONSISTENT WITH T-CONJUNCTIONS

C. Alsina
Secció de Matemàtiques i Informàtica,
Univ. Politècnica de Catalunya
Avda. Diagonal, 649. 08028 Barcelona
alsina@ea.upc.es

E. Trillas
Depto. Inteligencia Artificial,
Universidad Politécnica de Madrid
28660 Boadilla del Monte (Madrid)
etrillas@fi.upm.es

Summary

Motivated by some fundamental relations in Fuzzy Logic we study some basic cases of the functional equation $S_1(x,y) = S_2(x,T(N(x),y))$. In doing this some interesting techniques for solving functional equations are faced.

Key words: (S,N)-implications, t-norms, fuzzy connectives.

1 INTRODUCTION

In the context of Fuzzy Logic one considers frequently a triple of logical connectives *and, or* and *not* by introducing the functions (T,S,n) where T is a continuous t-norm, S is a continuous t-conorm and n is a strong negation. In this framework the modeling of fuzzy implications sometimes is made by means of the so-called S-implications, namely the functions

$$I(y|x) = S\left(n(x),y\right),$$

generalizing the classical identification of a conditional statement 'If p, then q' (represented by $p \rightarrow q$) with the statement 'Not p or q' represented by $\neg p \vee q$). When the classical setting is that of a Boolean Algebra B, the identity

$$p \rightarrow q = \neg p \vee q - \neg p \vee (p \wedge q) = p \rightarrow p \wedge q$$

enjoys the property $\neg p \wedge (p \wedge q) = 0$ that allows to some important laws of any probability P on B. For example, the proof that P is a non-decreasing function fro B to [0,1] **should** be obtained as follows: provided that $p \geq q$ (in the partial order of B) it is $q = p \vee q = p \vee (\neg p \wedge q)$ hence, in the linear order of $[0,1]$, it is $P(q) = P(p) + P(\neg p \wedge q) \geq P(p)$, because of the additivity of P. Consequently, it seems interesting to know when the identity $p \rightarrow q = p \rightarrow p \wedge q$ holds in a Fuzzy Logic on which the implication is modelled by an S-implication. In this context let's consider the following

Definition 1.1. Let $I_1(|)$ be a (S_1,n_1)-implication and let $I_2(|)$ be a (S_2,n_2)-implication. Fixed a conjunction T defined by a t-norm we will say that I_2 is T-**consistent** with I_1 whenever $I_2(T(x,y)|x) = I_1(y|x)$, i.e.,

$$S_2\left(n_2(x),T(x,y)\right) = S_1\left(n_1(x),y\right).$$

This equation corresponds to the fact that "$p \Rightarrow q$" and "$p \Rightarrow p \wedge q$" are conveniently identified.

Note that setting $y = 0$ yields $n_1 = n_2$. So we will consider only one negation $N = n_1 = n_2$ and we will take care of the condition

$$S_2\left(N(x),T(x,y)\right) = S_1\left(N(x),y\right). \qquad (1)$$

Next note that from (1) with $y = 1$ it follows that necessarily $S_2(N(x),x) = 1$, i.e., $I_2(x|x) = 1$ but (1) says nothing about the possible values of $I_1(x|x) = S_1\left(N(x),x\right)$.

Our aim in this paper is to solve completely (1) when $S_1(x,y) < 1$ whenever $x,y \neq 1$ and to study (1) also in some other related cases.

In doing this study we extend previous results found by us in [4] where equation (1) was considered in the case where N was precisely the strong negation associated to S_2. Here we solve (1) just with the restriction $S_1(x,y) < 1$ whenever $x,y \neq 1$. The other cases for S_1 will be treated elsewhere. In doing this we have developed some functional equations techniques that have their own mathematical interest. We assume notations and concepts as given in [3,9].

2 STUDY OF EQUATION (1)

Our aim here is to solve (1) when S_1 is a continuous t-conorm such that $S_1(x,y) < 1$ whenever, $x,y \neq 1$,

S_2 is a continuous t-conorm, T is a continuous t-norm and N is a strong negation. As we have seen before it follows from (1) that $S_2(N(x),x)=1$ so S_2 will be a non-strict Archimedean t-conorm representable in the form

$$S_2(x,y)=s_2^{(-1)}\left(s_2(x)+s_2(y)\right), \tag{2}$$

where $s_2:[0,1]\to[0,1]$ is continuous, strictly increasing, $s_2(0)=0$, $s_2(1)=1$ and $s_2^{(-1)}(x)=s_2^{-1}(x)$ for x in [0,1], $s_2^{(-1)}(x)=1$ for $x\geq 1$. Thus S_2 has its associated strong negation

$$N_2(x)=s_2^{-1}(1-s_2(x)), \tag{3}$$

and $S_2(x,y)=1$ if and only if $y\geq N_2(x)$. In particular since $S_2(x,N(x))=1$ we need to have $N\geq N_2$. Note that (1) is a functional equation of Pexider type with two variables x,y and four unknown functions S_2,S_1,T,N.

Lemma 2.1. Let S_2 and N_2 be given by (2) and (3), respectively. Let N be a strong negation and let T be a continuous t-norm satisfying

$$S_2\left(x,T(N(x),y)\right)=\mathrm{Max}(x,y), \tag{4}$$

for all x,y in [0,1]. This is possible if and only if T is a non-strict Archimedean t-norm with additive generator $t(x)=1-s_2(x)$ (i.e., T is the N_2-dual of $S_2: T=N_2\circ S_2\circ N_2\times N_2$) and $N=N_2=N_T$.

Lemma 2.2. If S_2 and N_2 are given by (2) and (3), respectively, N is a strong negation, T is a continuous t-norm, S_1 is a continuous t-conorm with $S_1(x,y)<1$ whenever $x,y\neq 1$ and (1) holds, then necessarily $N=N_2$ and T must have one of the following forms:

(i) $T=s_2^{-1}\circ\mathrm{Prod}\circ s_2\times s_2$

(ii) $T=s_2^{-1}\circ T_\lambda\circ s_2\times s_2$, where T_λ belongs to Frank's family.

(iii) $T=s_2^{(-1)}\circ W\circ s_2\times s_2$;

(iv) $T=$Min or T is an ordinal sum of Archimedean t-norms of the above type.

Proof. The commutativity of S_1 and (1) yield at once

$$\begin{aligned} S_1(x,y) &= S_2(x,T(N(x),y))=S_2(y,T(N(y),x)) \\ &= S_1(y,x), \end{aligned}$$

and since we are assuming $S_1(x,y)<1$ for $x,y\neq 1$ we deduce, using (2), that

$$s_2(x)+s_2(T(N(x),y))=s_2(y)+s_2(T(N(y),x)). \tag{5}$$

Introduce the new variables $u=s_2(y)$ and $v=s_2(N(x))$ into (6). Then for u,v in [0,1] arbitrary we obtain

$$\begin{aligned} &\left(s_2\circ N\circ s_2^{-1}\right)(v)+s_2\left(T(s_2^{-1}(v),s_2^{-1}(u))\right)= \\ &=u+\left(s_2\circ N\circ s_2^{-1}\circ s_2\circ N\right) \\ &\left(T(N(s_2^{-1}(u)),N(s_2^{-1}(v)))\right) \end{aligned} \tag{6}$$

Let $f:[0,1]\to[0,1]$ be defined by

$$f(t)=(s_2\circ N\circ s_2^{-1})(t). \tag{7}$$

Then f is strictly decreasing continuous and $f(0)=1$, $f(1)=0$. Consider now the t-norm:

$$G(u,v)=s_2\left(T(s_2^{-1}(u),s_2^{-1}(v))\right) \tag{8}$$

and the t-conorm

$$H(u,v)=(s_2\circ N)\left[T\left((N\circ s_2^{-1}(u),(N\circ s_2^{-1})(v)\right)\right]. \tag{9}$$

By virtue of the equation (7)

$$f(v)+G(u,v)=u+H(u,v),$$

therefore changing the roles of u,v we must have also

$$f(u)+G(u,v)=v+H(u,v),$$

whence substraction of the last equalities yield

$$f(v)-f(u)=u-v,$$

i.e.,

$$f(v)+v=f(u)+u,$$

and $f(x)+x$ is a constant function: $f(x)+x=k$. The condition $f(0)=1$ yields $k=1$, i.e., $f(x)=1-x$, whence by (8), $N(x)=s_2^{-1}(1-s_2(x))=N_2(x)$.

Now with $N=N_2$ we go back to (6) to get

$$\begin{aligned} &s_2(x)+s_2\left(T\left(s_2^{-1}(-s_2(x)),y\right)\right)= \\ &s_2(y)+s_2\left(T(s_2^{-1}(1-s_2(y),x)\right). \end{aligned} \tag{10}$$

The substitutions $a=s_2(x)$, $b=1-s_2(y)$ yield

$$\begin{aligned} &u+s_2\left[T(s_2^{-1}(1-u),s_2^{-1}(1-v))\right]= \\ &1-v+s_2\left[T(s_2^{-1}(v),s_2^{-1}(u))\right]. \end{aligned} \tag{11}$$

Thus the t-norm $T_2(u,v)=s_2\left[T(s_2^{-1}(u),s_2^{-1}(v)))\right]$ and its associated t-conorm

$$T_2^*(u,v)=1-T_2(1-u,1-v)$$

would satisfy by virtue of (12) Frank's equation

$$T_2(u,v)+T_2^*(u,v)=u+v,$$

i.e., T_2 is completely determined by Frank's theorem and so is T. The lemma follows.

Lemma 2.3. If S_2 and N_2 are given by (2) and (3), respectively, N is a strong negation, T is a continuous t-norm, S_1 is a strict t-conorm and (1) holds, then $N = N_2$ and there are two types or solutions either

(a) $T = s_2^{-1} \circ \text{Prod} \circ s_2 \times s_2$; $S_1 = s_2^{-1} \circ \text{Prod}^* \circ s_2 \times s_2$,

or

(b) $T = s_2^{-1} \circ T_\lambda \circ s_2 \times s_2$; $S_1 = s_2^{-1} \circ T^*_{1/\lambda} \circ s_2 \times s_2$;

where T_λ belongs to Frank's family (see Theorem 2.2.).

Proof. If S_1 is a strict t-conorm, $S_1(x,y) < 1$ whenever $x \cdot y \neq 1$ and we can apply Lemma 2.2, so $N = N_2$ and the possibles forms of the t-norm T are determined. Then checking which operations one can be consistent with the fact that (1) holds for a strict t-conorm S_1 we obtain the lemma.

Lemma 2.4. If S_2 and N_2 are given by (2) and (3), respectively, T is a continuous t-norm, N is a strong negation, S_1 is an ordinal sum of Archimedean t-conorms and (1) holds then $S_2(x,y) < 1$ for all $x, y \neq 1$.

Proof. Under the above conditions consider the set $O = \{(x,y) \in [0,1]^2 \mid S_1(x,y) = 1\}$. Since S_1 is an ordinal sum, if there would exist a point $(x_0, y_0) \in O$ with $x_0, y_0 \neq 1$ then S, would take the value 1 in a region of the form

$$P = \{(x,y) \mid x \in [b,1],\ y \geq g(x)\} \subset O,$$

for some b in (0,1) idempotent element of S_1 and for some function $g : [b,1] \to [b,1]$ continuous, strictly decreasing, $g(b) = 1$, $g(1) = b$ and $g = g^{-1}$.

When (x,y) moves in P the corresponding point $(N(x), y)$ moves in

$$Q = \{(a,b) \mid 0 \leq a \leq N(b),\ b \geq g(N(a))\}$$

and in this case $(x,y) \in P$ and (1) imply $1 = S_1(x,y) = S_2(x, T(N(x), y))$, i.e., $T(N(x), y) \geq N_2(x)$, inquality which implies in particular that T can not vanish in Q.

Next note that we must have S_1 =Max on $[b,1] \times [0,b] \cup [0,b] \times [b,1]$ whence for $0 \leq y \leq b \leq x \leq 1$

$$x = \text{Max}(x,y) = S_1(x,y) = S_2(x, T(N(x), y)),$$

and necessarily $T(N(x), y) = 0$, but when (x,y) moves in $[b,1] \times [0,b]$ the points $(N(x), y))$ move in $K = [0, N(b)] \times [0,b]$, i.e., $T \equiv 0$ on K. Recalling the representation theorem, T can not be Min or a strict t-norm because T vanishes on K, T can not be a non-strict Archimedean t-norm because T does not vanish in Q,... and T can not be an ordinal sum because its vanishing Archimedean component would need to include Q in its zero set and therefore its zero set would need to cut Q efectively, which is not possible. We get a contradiction and the lemma is proved. □

Lemma 2.5. If S_2 and N_2 are given by (2) and (3), respectively, N is a strong negation and S_1 is an ordinal sum of Archimedean t-norms then there exists no continuous t-norm satisfying (1).

From all the above lemmata it follows at once our main result

Theorem 2.1. Let S_2 be a non-strict Archimedean t-conorm with additive generator s_2 and let N_2 be its associated strong negation (3). Let N be a strong negation. Then a continuous t-norm T and a continuous t-conorm S_1 such that $S_1(x,y) < 1$ for $x, y \neq 1$, satisfy (1) if and only if one of the following conditions hold:

(i) S_1 =Max, $T = N_2 \circ S_2 \circ N_2 \times N_2$, $N = N_2 = N_T$;

(ii) $S_1 = s_2^{-1} \circ \text{Prod}^* \circ s_2 \times s_2$, $T = s_2^{-1} \circ \text{Prod} \circ s_2 \times s_2$, $N = N_2$;

(iii) $S_1 = s_2^{-1} \circ T^*_{1/\lambda} \circ s_2 \times s_2$, $T = s_2^{-1} \circ T_\lambda \circ s_2 \times s_2$; $N = N_2$;

References

[1] J. Aczél. Lectures on Fucntional Equations and Their Applications, Academic Press, New York, 1966.

[2] C. Alsina. On connectives in Fuzzy Logic Satisfying the Condition $S(T_1(x,y), T_2(x, N(y))) = x$. Proc. FUZZ'IEEE-97, Barcelona 1997, 149-153.

[3] C. Alsina. As you like them: connectives in Fuzzy Logic, ISMVL 96, Santiago de Compostela, 1996, Proc. ISMVL, 1-7.

[4] C. Alsina and E. Trillas. When (S,N)-implications are (T,T_1)-conditional functions? 1998 (submitted).

[5] C. Alsina, M.J. Frank and B. Schweizer. *A primer of t-norms. Associative functions on real intervals*, (1999).

[6] C. Alsina, E. Trillas and L. Valverde. On some logical connectives for Fuzzy Set Theory, *Journal of Mathematical Analysis and Applications*, 93, 1 (1983), pp. 15-26.

[7] G. Birkhoff. *Lattice theory*, Pub. A.M.S., Providence, 1971.

[8] J.C. Fodor and M. Roubens. *Fuzzy preference modelling and multicriteria decision support*, Kluwer, Dordrecht, 1994.

[9] B. Schweizer and A. Sklar. *Probabilistic Metric Spaces*, Elsevier North-Holland, New York (1983).

[10] E. Trillas. Sobre funciones de negación en la teoría de los subconjuntos difusos" (in spanish), *Stochastica* (1979) III-1, pp. 47-60. Reprinted (english version) in *Advances in Fuzzy Logic*, Barro, S. et altri ed. Publicacións Universidade de Santiago de Compostela (1998) pp. 31-43.

[11] E. Trillas and L. Valverde. "On Implication and Indistinguishability in the setting of Fuzzy Logic" in R.R. Yager and J. Kacpryck Eds. *Management Decision Support System using Fuzzy Set and Possibility Theory* (1985), North-Holland pp 198-212.

[12] E. Trillas, C. Alsina and J. Jacas. On logical connectives for a fuzzy set theory with or without non-empty self-contraditions, 1998 (submitted).

[13] E. Trillas, C. Alsina and J.M. Terricabras. *Introducción a la lógica borrosa*, Ariel, Barcelona, (1995), (in spanish).

[14] L.A. Zadeh. "Fuzzy Sets", *Information and Control*, 8 (1965) pp. 33-50.

[15] L.A. Zadeh. "Fuzzy Logic and Approximate Reasoning", *Synthese*, 30 (1975) pp. 407-428.

Fuzzy projective geometries

Leen Kuijken*

Department of Applied Mathematics and Computer Science
University of Ghent, Krijgslaan 281 (S9),
B-9000 Gent, BELGIUM
lk@cage.rug.ac.be

Summary

In this paper we introduce a new model of fuzzy projective geometries. This model is richer than the one introduced in [4] and [5], since points and lines play a similar role, like they do in classical projective geometry. Furthermore, we will show that this new fuzzy projective geometry is closely related to the fibred projective geometries introduced in [6].
Keywords: Fuzzy geometry, Fibred geometry.

1 Introduction

In [4] and [5] we introduced a first model of fuzzy projective geometries, deduced from respectively fuzzy vector spaces and fuzzy groups. This provided a link between the fuzzy versions of classical theories that are very closely related. However, the geometric structure involved in this model is rather weak: a fuzzy projective space in this sense is equivalent with a given sequence of subspaces in the base projective space.
In [6] another fuzzy model of projective geometries was constructed: fibred projective planes. In this model the role of points and lines is equivalent (this is not the case in the first model), as in the classical case. Points and lines of the base geometry mostly have multiple degrees of membership.
This paper introduces a third model. We first define a fuzzy projective plane in which points and lines play the same role, and such that every point and every line in the base plane possess only one degree of membership. Afterwards we give a definition for an n-dimensional fuzzy projective space. We also investigate the link between fibred and fuzzy projective geometries.

*The author is a Research Fellow supported by the Flemish Institute for the Promotion of Scientific and Technological Research in Industry (IWT), grant No. 961336

2 Preliminaries

Definition 2.1 ([2])
An **(axiomatic) projective plane** $\mathcal{P}$ is an incidence structure (P, B, I) with P a set of points, B a set of lines and I an incidence relation, such that the following axioms are satisfied:
(A1) every pair of distinct points are incident with a unique common line;
(A2) every pair of distinct lines are incident with a unique common point;
(A3) $\mathcal{P}$ contains a set of four points with the property that no three of them are incident with a common line.
A **closed configuration** S of $\mathcal{P}$ is a subset of $P \cup B$ that is closed under taking intersection points of any pair of lines in S and lines spanned by any pair of distinct points of S. We denote the line in $\mathcal{P}$ spanned by the points a and b by $\langle a, b\rangle$.

Definition 2.2 ([1], p. 10)
An **(axiomatic) projective space** $\mathcal{S}$ is an incidence structure (P, B, I) with P a set of points, B a set of lines and I an incidence relation, such that the following axioms are satisfied:
(A1) every line is incident with at least two points;
(A2) every pair of distinct points are incident with a unique common line;
(A3) given distinct points a, b, c, d, e such that $\langle a, b\rangle = \langle a, c\rangle \neq \langle a, d\rangle = \langle a, e\rangle$, there is a point $xI\langle b, d\rangle \cap \langle c, e\rangle$ (Pasch's axiom).

Definition 2.3 ([7])
A **fuzzy set** λ on a set X is a mapping $\lambda : X \to [0, 1]$:

$x \mapsto \lambda(x)$. The number $\lambda(x)$ is called the **degree of membership** of the point x in λ.

We denote the minimum operator by $\wedge$.

Definition 2.4 ([4], [5])
A fuzzy set λ on the n-dimensional projective space $\mathcal{S}$ is a **fuzzy n- dimensional projective space** on $\mathcal{S}$ if $\lambda(p) \geq \lambda(q) \wedge \lambda(r)$, for any three collinear points p, q, r of $\mathcal{S}$. We denote $[\lambda, \mathcal{S}]$.
The projective space $\mathcal{S}$ is called the **base projective space** of $[\lambda, \mathcal{S}]$. If $\mathcal{S}$ is a fuzzy point, line, plane, ..., we use base point, base line, base plane, ..., respectively.

Definition 2.5 ([6])
Consider the projective plane $\mathcal{P} = (P, B, I)$. Suppose $a \in P$ and $\alpha \in]0, 1]$. The **f-point** (a, α) is the following fuzzy set on the point set P of $\mathcal{P}$:

$$\begin{array}{lccll} (a,\alpha): & P & \to & [0,1] & \\ & a & \mapsto & \alpha & \\ & x & \mapsto & 0 & \text{if } x \in P \backslash \{a\}. \end{array}$$

The point a is called the **base point** of the f-point (a, α). An **f-line** (L, α) with base line L is defined in a similar way.

With the definition of the intersection of fuzzy sets (see [7]) and the extension principle (see [3]), we see that:

Definition 2.6 ([6])
The f-lines (L, α) and (M, β) intersect in the unique f-point $(L \cap M, \alpha \wedge \beta)$.
The f-points (a, λ) and (b, β) span the unique f-line $(\langle a, b \rangle, \lambda \wedge \beta)$.

Definition 2.7 ([6])
A **fibred projective plane** $\mathcal{FP}$ on the projective plane $\mathcal{P}$ consist of a set $\mathcal{F}P$ of f-points and a set $\mathcal{F}B$ of f-lines, such that every point and line of $\mathcal{P}$ is base point and base line of at least one f-point and f-line respectively, and such that $(\mathcal{F}P, \mathcal{F}B)$ satisfies the following fuzzified axioms of a projective plane (cf. definition 2.1):
(F1) every pair of f-points with distinct base points span a unique f-line;
(F2) every pair of f-lines with distinct base lines intersect in a unique f-point;
The projective plane $\mathcal{P}$ is called the **base geometry** of $\mathcal{FP}$.

We can construct a fibred projective plane in the following way (see [6]). Let $P' \subseteq P$ and $B' \subseteq B$ be such that the unique closed configuration containing $P' \cup B'$ is $P \cup B$. For each element x of $P' \cup B'$, we choose arbitrarily a nonempty subset Σ_x of $]0, 1]$ of which the elements are called the **initial values of** x, and we define a fibred projective plane $\mathcal{FP}$ as follows. For each $x \in P' \cup B'$ and for each $\alpha \in \Sigma_x$, the element (x, α) belongs to $\mathcal{FP}$. This is step 1 of the construction. We now describe step i, $i > 1$.

For any pair of f-points that we already obtained, the f-line spanned by it also belongs to $\mathcal{FP}$ by definition. Dually, for any pair of f-lines, the intersection f-point belongs to $\mathcal{FP}$. The set of all f-points and of all f-lines constructed this way in a finite number of steps is readily verified to constitute a fibred projective plane.

It is clear that every fibred projective plane can be constructed as above. Indeed, one can always take for each element all its corresponding values as initial values.

Now suppose Σ_x is a singleton for every $x \in P \cup B$. If $P' = P$ and $B' = \emptyset$, then we call the fibred projective plane **mono-point-generated**. If $P' = P$ and $B' = B$, then the fibred projective plane is called **mono-generated**. We will restrict ourselves to these two kinds of fibred projective planes.

We see that $\mathcal{FP}$ can be considered as an ordinary projective plane (its base plane $\mathcal{P}$) where to every point and line, a set of values from $]0, 1]$ are assigned. Also the fuzzy projective plane in definition 2.4 can be considered as an ordinary projective plane, where to every point (and only to points) one (and only one) degree of membership is assigned.

Example: (see [6])
Consider the classical projective plane $\mathcal{F} = GF(2, 2)$, the Fano plane. We will construct a mono-point-generated fibred projective plane with base plane $\mathcal{F}$. We label the 7 points of $\mathcal{F}$ as $\{a, b, c, d, e, f, g\}$ and the lines as $\{A, B, C, D, E, F, G\}$, such that:
$A = \{a, b, c\}$,
$B = \{c, d, e\}$,
$C = \{e, f, a\}$,
$D = \{a, g, d\}$,
$E = \{b, g, e\}$,
$F = \{c, g, f\}$,
$G = \{b, d, f\}$.

In step 1, we construct the f-points $(a, 0.9), (b, 0.8), (c, 0.7), (d, 0.6), (e, 0.3), (f, 0.4)$ and $(g, 0.5)$ on the points of P, thus 0.9, 0.8, 0.7, 0.6, 0.3, 0.4, and 0.5 are the initial values of the resp. base points a, b, c, d, e, f and g. Following the foregoing construction, these initial values yield the following fibred

projective plane:

$\Sigma_a = \{0.3, 0.4, 0.5, 0.6, 0.9\}$,
$\Sigma_b = \{0.3, 0.4, 0.5, 0.6, 0.8\}$,
$\Sigma_c = \{0.3, 0.4, 0.5, 0.6, 0.7\}$,
$\Sigma_d = \{0.3, 0.4, 0.5, 0.6\}$,
$\Sigma_e = \{0.3, 0.4, 0.5\}$,
$\Sigma_f = \{0.3, 0.4, 0.5\}$,
$\Sigma_g = \{0.3, 0.4, 0.5\}$, and

$\Sigma_A = \{0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$,
$\Sigma_B = \{0.3, 0.4, 0.5, 0.6\}$,
$\Sigma_C = \{0.3, 0.4, 0.5\}$,
$\Sigma_D = \{0.3, 0.4, 0.5, 0.6\}$,
$\Sigma_E = \{0.3, 0.4, 0.5\}$,
$\Sigma_F = \{0.3, 0.4, 0.5\}$,
$\Sigma_G = \{0.3, 0.4, 0.5, 0.6\}$.

3 Fuzzy projective planes

In this section we introduce a third model of a fuzzy projective geometries. Like in the fibred model, it also assigns values to the lines of the base geometry. Like the model in definition 2.4 it assigns only *one* value to every point (and line) of the base geometry, as custom in most definitions in the area of fuzzy algebra. In the sequel, a fuzzy projective plane will refer to the following definition, no longer to definition 2.4, so there will be no confusion.

Definition 3.1 Suppose $\mathcal{P}$ is a projective plane (P, B, I). The fuzzy set μ on $P \cup B$ is a **fuzzy projective plane** on $\mathcal{P}$ if
(1) $\mu(L) \geq \mu(p) \wedge \mu(q)$, $\forall p, q : \langle p, q \rangle = L$ and
(2) $\mu(p) \geq \mu(L) \wedge \mu(M)$, $\forall L, M : L \cap M = p$.

Definition 3.2 Consider the fibred projective plane $\mathcal{FP}$ with base plane the projective plane $\mathcal{P} = (P, B, I)$. As in section 2, let Σ_p (resp. Σ_L) be the set of all different degrees of membership of a point p (resp. line L), for all $p \in P$ and $L \in B$. **Skimming** $\mathcal{FP}$ means that for every element x of $\mathcal{P}$, we only keep the highest degree of membership, i.e. $\sup \Sigma_x$. This results in a fuzzy set χ on the base projective plane $\mathcal{P}$, called the **cream** of the fibred projective plane $\mathcal{FP}$. Thus:

$$\begin{array}{rccl} \chi: & P \cup B & \rightarrow & [0,1] \\ & x & \mapsto & \sup \Sigma_x \end{array}$$

Theorem 3.1 *The cream of a fibred projective plane is a fuzzy projective plane, and every fuzzy projective plane can be considered as the cream of a fibred projective plane.*

This theorem makes sure the new definition makes sense: since fibred projective planes exist, fuzzy projective planes will also exist.

Example:
By the previous theorem, we know that the example in section 2 gives rise to the following fuzzy projective plane μ on the Fano plane $\mathcal{F}$:
$\mu(a) = 0.9$, $\mu(b) = 0.8$, $\mu(c) = 0.7$, $\mu(d) = 0.6$, $\mu(e) = \mu(f) = \mu(g) = 0.5$ and
$\mu(A) = 0.8$, $\mu(B) = 0.6$, $\mu(C) = 0.5$, $\mu(D) = 0.6$, $\mu(E) = \mu(F) = 0.5$, $\mu(G) = 0.6$.

4 Fuzzy projective spaces

So far we have only considered 2-dimensional fibred and fuzzy projective geometries, i.e. the plane case. We can also define n-dimensional fibred and fuzzy projective geometries, with n an arbitrary finite integer, such that the previous theorem holds in the general case. Consider the n-dimensional projective space $\mathcal{S}$. Call U_i the set of all i-dimensional subspaces of $\mathcal{S}$, for all i: $0 \leq i \leq n-1$.

Definition 4.1 Suppose $\mathcal{S}$ is an n-dimensional projective space, and $i < n$. An **f-subspace** (V_i, α) **of dimension** i is the following fuzzy set on the set U_i

$$\begin{array}{rccll} (V_i, \alpha): & U_i & \rightarrow & [0,1] & \\ \text{of } \mathcal{S}: & V_i & \mapsto & \alpha & \\ & x & \mapsto & 0 & \text{if } x \in U_i \backslash \{V_i\}. \end{array}$$

The subspace V_i is the **base subspace** of (V_i, α).

Definition 4.2 An n-dimensional **fibred projective space** $\mathcal{FS}$ on the n-dimensional projective space $\mathcal{S}$ consist of n sets of f-objects: f-subspaces of dimension i, for $0 \leq i \leq n-1$. Every subspace (of dimension i) of $\mathcal{S}$ is base subspace of at least one f-subspace (of dimension i). Moreover the following axioms have to be fullfilled:

(F1) the intersection of two f-subspaces (with distinct base subspaces that are not disjoint) is again an f-subspace;

(F2) every two f-subspaces (with distinct base subspaces that do not span $\mathcal{S}$ itself) span an f-subspace.

For $i = 0, 1, 2, n-1$, the f-subspaces of dimension i will be called f-points, f-lines, f-planes and f-hyperplanes.

The **cream** of an n-dimensional fibred projective space is defined in the same way as for a fibred projective plane (see definition 3.2)

Definition 4.3 Suppose $\mathcal{S}$ is an n-dimensional projective space as defined above. The fuzzy set μ on $\bigcup_{i=0}^{n-1} U_i$ is a **fuzzy projective space** of dimension n on $\mathcal{S}$ if for all subspaces V_i, V_j, V_k, $0 \leq i, j, k \leq n-1$ we have:
(1) $\mu(V_i) \geq \mu(V_j) \wedge \mu(V_k)$, $\forall\, V_j$ and V_k, $V_j \neq V_k$ such that $V_j \cap V_k = V_i$ if $V_i \neq \emptyset$ and
(2) $\mu(V_i) \geq \mu(V_j) \wedge \mu(V_k)$, $\forall\, V_j$ and V_k, $V_j \neq V_k$ such that $\langle V_j, V_k \rangle = V_i$ if $V_i \neq \mathcal{S}$.

Theorem 4.1 *The cream of an n-dimensional fibred projective space is an n-dimensional fuzzy projective space, and every n-dimensional fuzzy projective space can be considered as the cream of an n-dimensional fibred projective space.*

Acknowledgment
This results are part of the research done for my PhD thesis, supervised by Prof. H. Van Maldeghem and Prof. E. Kerre.

References

[1] F. Buekenhout, *Handbook of Incidence Geometry, Buildings and Foundations*, North – Holland, Amsterdam, 1995.

[2] D. R. Hughes and F. C. Piper, *Projective Planes*, Springer – Verlag, New York Heidelberg Berlin, 1973.

[3] E. E. Kerre (Ed.), *Introduction to the basic principles of fuzzy set theory and some of its applications*, Communication & Cognition, Gent/Belgium, 1993.

[4] L. Kuijken, H. Van Maldeghem and E. Kerre, *Fuzzy Projective Geometries from Fuzzy Vector Spaces*, Proceedings of IPMU '98, 1331 – 1338.

[5] L. Kuijken, H. Van Maldeghem and E. Kerre, *Fuzzy Projective Geometries from Fuzzy Groups*, Tatra Mountains Mathematical Publications, to appear.

[6] L. Kuijken, H. Van Maldeghem, *Fibred geometries*, Discrete Mathematics, submitted for publication.

[7] L. Zadeh, Fuzzy Sets, *Information and control* **8** (1965), 338–358.

PROPERTIES OF THE FUZZY COMPOSITIONS BASED ON AGGREGATION OPERATORS AND ITS PSEUDOCOMPLEMENT

Miren Portilla
Dpto. Estadística e Inv. Operativa
Universidad Pública de Navarra
31006 Pamplona, Spain
e-mail: mportilla@si.upna.es

Pedro Burillo
Dpto. Automática y Computación
Universidad Pública de Navarra
31006 Pamplona, Spain
e-mail: pburillo@si.upna.es

Abstract

The main aim of this paper is to present the properties of the $sup - F$ composition based on the aggregation operator F defined by Mayor and Torrens and analyze some properties about the pseudocomplement φ and the $inf - \varphi$ composition, to show that these compositions have analogous properties to the based on t-norms.

Keywords: Aggregation operators, composition of fuzzy relations, t-norms, t-relative pseudocomplement.

1 INTRODUCTION

The problem of finding appropriate connectives for the logical combination of fuzzy sets and fuzzy relations has turned to be an important issue from several points of view. In this context, triangular norms and conorms have shown to be useful tools in multiple fields, and more recently the aggregation operators play also an important role in fuzzy sets theory and in applied areas as social choice, multicriteria decision making, synthesis of implication functions, etc. But apart from the applications of these aggregation operators in the fields above mentioned, there is an increasing interest in their theoric study, which no doubt is going to lead to new possibilities of applications.

The aggregation has been defined by many authors as fuzzy operators generalizing "*and*" and "*or*" fuzzy connectives ([8], [11], [12], [14], [16], [22]) and has been applied as a consensus of fuzzy opinions in several contexts ([5], [6], [9], [10], [12], [13], [17], [19], [23]). Moreover nearly each author has proposed his own definition. Anyway, we can state that the axioms for the definition of these aggregation operators F, defined in the unit interval, whether binary or n-ary with $n \geq 3$, usually requires that F must be commutative, non-decreasing in each place and verify the boundary conditions $F(0,...,0) = 0$ and $F(1,...,1) = 1$. In this respect, the binary aggregations of Cholewa [6], Dombi [8], Fung & Fu [11], Mizumoto [16] and Zimmermann & Zysno [23], and the n-ary aggregations of Dubois & Koning [9], Fodor & Roubens [10], Klir & Folger [12], Montero [17], Ramakrishnan & Rao [19] and Yager [22] verify these conditions.

One of the aggregation operators which verify these properties has been introduced by Mayor & Torrens [14] as a generalization of t-norms and t-conorms, and it is given in the next definition.

Definition 1. *An aggregation operator is a commutative binary operation F in $[0,1]$, non-decreasing in each place, such that*

$$F(x,0) = kx \qquad F(x,1) = (1-k)x + k$$

for every $x \in [0,1]$, where $k = F(0,1)$.

This operator includes both t-norms and t-conorms and even any convex linear combination of a t-norm and a t-conorm. In this respect, Alsina, Mayor, Tomas & Torrens have proved in [1] that F is a t-norm if and only if it is associative and $k = 0$, and also that F is a t-conorm if and only if it is associative and $k = 1$. Moreover, we have that F is associative if and only if it is a t-norm (and then $k = 0$) or a t-conorm (and then $k = 1$).

In this context, we study in the next section the $sup - F$ composition based in the aggregation operator, as a generalization of $sup - t$ composition based on t-norms t. Moreover, the main aim of the sections 3 and 4 is to define the pseudocomplement φ associated with the aggregation operator and the $inf - \varphi$ composition and analyze their properties, as an extension of the t-relative pseudocomplement analized by Weber [21], Miyakoshi and Shimbo [15] and Klir and Yuan [12].

2 SUP-F COMPOSITION OF FUZZY RELATIONS

It is well known that the $sup - t$ composition of binary fuzzy relations, where t refers to a t-norm, generalize the standard $max - min$ composition and the $max - *$ composition (analyzed by Bezdek & Harris [2] and Rosenfeld [20]). In the same way, from the above binary aggregation F, we can define the $sup - F$ composition of two fuzzy binary relations as we defined in [4] and [3]. This $sup - F$ composition generalize also the standard $max - min$ composition, the $max - *$ composition and even the $sup - t$ composition.

Let X, Y be two nonempty sets and $[0,1]$ be the lattice with the usual "$\vee$" and "$\wedge$" operators. If we denote by $[0,1]^{X\times Y}$ the set of all the fuzzy binary relations, i.e., the fuzzy sets in the product space $X \times Y$, then we can define the $sup - F$ composition as follows:

Definition 2. *The $sup - F$ composition of two fuzzy binary relations $P \in [0,1]^{X\times Y}$ and $Q \in [0,1]^{Y\times Z}$, is the fuzzy relation given by:*

$$(P \underset{\vee}{@} Q)(x,z) = \bigvee_{y\in Y} F(P(x,y), Q(y,z))$$

for all $x \in X$, $z \in Z$.

Now we present the properties of the previous composition analyzed in [18], in order to show that they verify analogous properties to the $sup - t$ composition [12].

The basic properties of the $sup - F$ composition under standard fuzzy union and intersection, that follow directly from commutativity and monotonicity of the aggregation F, are the following:

$$(\bigcup_{i\in I} P_i) \underset{\vee}{@} Q = \bigcup_{i\in I} (P_i \underset{\vee}{@} Q) \tag{1}$$

$$P \underset{\vee}{@} (\bigcup_{i\in I} Q_i) = \bigcup_{i\in I} (P \underset{\vee}{@} Q_i) \tag{2}$$

$$(\bigcap_{i\in I} P_i) \underset{\vee}{@} Q \subseteq \bigcap_{i\in I} (P_i \underset{\vee}{@} Q) \tag{3}$$

$$P \underset{\vee}{@} (\bigcap_{i\in I} Q_i) \subseteq \bigcap_{i\in I} (P \underset{\vee}{@} Q_i) \tag{4}$$

$$(P \underset{\vee}{@} Q)^{-1} = Q^{-1} \underset{\vee}{@} P^{-1} \tag{5}$$

for any fuzzy relations $P, P_i \in [0,1]^{X\times Y}$ and $Q, Q_i \in [0,1]^{Y\times Z}$, where i takes values in a finite index set I. However, this composition are obviously not commutative.

Remark: The study of the basic properties that remain valid for arbitrary index sets is not included because the aim of this paper is to present the main properties of this composition. However, we are analyzing this topic.

The $sup - F$ composition is also monotonic non-decreasing, that is, for any fuzzy relations $P \in [0,1]^{X\times Y}$, $Q_1, Q_2 \in [0,1]^{Y\times Z}$ and $R \in [0,1]^{Z\times T}$, if $Q_1 \subseteq Q_2$ then

$$P \underset{\vee}{@} Q_1 \subseteq P \underset{\vee}{@} Q_2 \qquad Q_1 \underset{\vee}{@} R \subseteq Q_2 \underset{\vee}{@} R \tag{6}$$

In relation to the associativity of the $sup - F$ composition, we have the following propositions [18]:
1) If the universal sets Y and Z are finite, then $[P \underset{\vee}{@} Q] \underset{\vee}{@} R = P \underset{\vee}{@} [Q \underset{\vee}{@} R]$ for any $P \in [0,1]^{X\times Y}$, $Q \in [0,1]^{Y\times Z}$ and $R \in [0,1]^{Z\times T}$, if and only if the aggregation is associative.
2) If the universal sets Y and Z are not finite, then $[P \underset{\vee}{@} Q] \underset{\vee}{@} R = P \underset{\vee}{@} [Q \underset{\vee}{@} R]$ for any $P \in [0,1]^{X\times Y}$, $Q \in [0,1]^{Y\times Z}$ and $R \in [0,1]^{Z\times T}$, if and only if the aggregation is associative and lower semicontinuous.

As we have said, the aggregation F is associative if and only if is a t-norm or a t-conorm, then t-norms and t-conorms are the only aggregations of Mayor and Torrens that verify these last two propositions. That confirms that the $sup - F$ composition generalizes the $sup - t$ composition and have anologous properties.

3 PSEUDOCOMPLEMENT ASSOCIATED WITH AN AGGREGATION

Weber [21] introduced in 1983 the concept of t-relative pseudocomplement associated with a continuous t-norm, also called residual operator by Fodor and Roubens [10]. This operator have been applied in the resolution of fuzzy relation equations with composition based on t-norms by Miyakoshi and Shimbo in [15] and Di Nola et al. in [7].

In this section we are going to present the pseudocomplement associated with an aggregation lower semicontinuous, as a generalization of the t-relative pseudocomplement of a t-norm. Moreover, we analyze its properties to show that their are analogous to the properties of the t-relative pseudocomplement analyzed by Miyakoshi and Shimbo in [15] and Klir and Yuan in [12].

Definition 3. *Given a lower semicontinuous aggregation F of Mayor and Torrens with $k = F(0,1) = 0$, its pseudocomplement φ is given by*

$$a \varphi b = sup\{x \in [0,1] : F(a,x) \leq b\}$$

for every $a, b \in [0,1]$.

We note that the aggregation operator F must verify $k = F(0,1) = 0$ in order to let the set $\{x \in [0,1] : F(a,x) \leq b\}$ be non-empty, but only the lower semicontinuity of F is needed.

As we have said, this operator generalizes the defined with a continuous t-norm and plays an important role in fuzzy relation equations ([15], [12]). While the aggregation F may be interpreted as logical conjunction, the corresponding operation φ may be interpreted as logical implication. Basic properties of φ are expressed by the following theorems.

Theorem 1. *For any $a, a_1, a_2, b, b_1, b_2, c \in [0,1]$, operation φ verifies:*
(p-1) $c \leq a\varphi b$ iff $F(a,c) \leq b$
(p-2) $(a\varphi b)\varphi b \geq a$
(p-3) $b_1 \leq b_2$ implies $a\varphi b_1 \leq a\varphi b_2$
(p-4) $a_1 \leq a_2$ implies $a_1\varphi b \geq a_2\varphi b$
(p-5) $F(a, a\varphi b) \leq b$
(p-6) $a\varphi F(a,b) \geq b$

Theorem 2. *For any $a, a_i, b, b_i \in [0,1]$, where i takes values from and index set, operation φ verifies:*
(q-1) $(\bigvee_{i\in I} a_i)\varphi b = \bigwedge_{i\in I}(a_i\varphi b)$
(q-2) $(\bigwedge_{i\in I} a_i)\varphi b \geq \bigvee_{i\in I}(a_i\varphi b)$
(q-3) $a\varphi(\bigwedge_{i\in I} b_i) = \bigwedge_{i\in I}(a\varphi b_i)$
(q-4) $a\varphi(\bigvee_{i\in I} b_i) \geq \bigvee_{i\in I}(a\varphi b_i)$

4 INF-φ COMPOSITION OF FUZZY RELATIONS

Given an aggregation operator F of Mayor and Torrens and the associated operator φ, the $inf - \varphi$ composition of two fuzzy relations is given in the next definition.

Definition 4. *The $inf - \varphi$ composition of two fuzzy binary relations $P \in [0,1]^{X\times Y}$ and $Q \in [0,1]^{Y\times Z}$, is the fuzzy relation given by:*

$$(P \underset{\wedge}{\Phi} Q)(x,z) = \bigwedge_{y\in Y}(P(x,y)\varphi Q(y,z))$$

for all $x \in X$, $z \in Z$.

Basic properties of the $inf - \varphi$ composition under standard fuzzy union and intersection, that follow directly from theorem 2, are the following:

$$(\bigcup_{i\in I} P_i)\underset{\wedge}{\Phi} Q = \bigcap_{i\in I}(P_i \underset{\wedge}{\Phi} Q) \tag{7}$$

$$(\bigcap_{i\in I} P_i)\underset{\wedge}{\Phi} Q \supseteq \bigcup_{i\in I}(P_i \underset{\wedge}{\Phi} Q) \tag{8}$$

$$P \underset{\wedge}{\Phi}(\bigcap_{i\in I} Q_i) = \bigcap_{i\in I}(P \underset{\wedge}{\Phi} Q_i) \tag{9}$$

$$P \underset{\wedge}{\Phi}(\bigcup_{i\in I} Q_i) \supseteq \bigcup_{i\in I}(P \underset{\wedge}{\Phi} Q_i) \tag{10}$$

for any fuzzy relations $P, P_i \in [0,1]^{X\times Y}$ and $Q, Q_i \in [0,1]^{Y\times Z}$, where i takes values in a finite index set I.

The previous properties allows us to prove two properties of monotonicity of the $inf - \varphi$ composition: for any fuzzy relations $P \in [0,1]^{X\times Y}$, $Q_1, Q_2 \in [0,1]^{Y\times Z}$ and $R \in [0,1]^{Z\times T}$, if $Q_1 \subseteq Q_2$ then

$$P \underset{\wedge}{\Phi} Q_1 \subseteq P \underset{\wedge}{\Phi} Q_2 \tag{11}$$

$$Q_1 \underset{\wedge}{\Phi} R \supseteq Q_2 \underset{\wedge}{\Phi} R \tag{12}$$

These properties enable us to prove two theorems which play an important role in the resolution of fuzzy relation equations based on $sup - F$ and $inf - \varphi$ compositions ([7], [12]), since they provide us a procedure to determine the maximum and minimum solutions of these fuzzy relation equations.

Theorem 3. *Let $P \in [0,1]^{X\times Y}$, $Q \in [0,1]^{Y\times Z}$ and $R \in [0,1]^{Z\times T}$ be fuzzy relations. Then,*

$$P \underset{\vee}{@} Q \subseteq R \quad iff \quad Q \subseteq P^{-1} \underset{\wedge}{\Phi} R \tag{13}$$

Theorem 4. *Let $P \in [0,1]^{X\times Y}$, $Q \in [0,1]^{Y\times Z}$ and $R \in [0,1]^{Z\times T}$ be fuzzy relations. Then,*

$$P^{-1} \underset{\vee}{@}(P \underset{\wedge}{\Phi} Q) \subseteq Q \tag{14}$$

$$R \subseteq P \underset{\wedge}{\Phi}(P^{-1} \underset{\vee}{@} R) \tag{15}$$

$$P \subseteq (P \underset{\wedge}{\Phi} Q) \underset{\wedge}{\Phi} Q^{-1} \tag{16}$$

$$R \subseteq (R \underset{\wedge}{\Phi} Q^{-1}) \underset{\wedge}{\Phi} Q \tag{17}$$

The previous properties and theorems have been proved by Klir and Yuan in [12] for the $inf - w_t$ composition based on the operator w_t associated with a continuous t-norm t. Therefore, the $inf - \varphi$ composition generalizes the $inf - w_t$ composition, as we wanted to show.

Conclusions

In this paper we have proved that the $sup - F$ and $inf - \varphi$ compositions of fuzzy relations based on the aggregation operator F of Mayor and Torrens and its pseudocomplement φ, respectively, generalize the $sup - t$ and $inf - w_t$ compositions based on the t-norm t and its pseudocomplement w_t, since they have analogous properties. This analysis is the first step to analyze the solutions of fuzzy relation equations based on these compositions.

In this way, our investigation will continue analyzing the fuzzy relation equations with $sup - F$ and $inf - \varphi$ compositions, for any aggregation operator F and its pseudocomplement φ, and also applying these results in fuzzy decision making, because we think that these properties could be interesting for analysis with fuzzy preference relations and consensus of fuzzy opinions.

Acknowledgements

The authors are grateful to the referees for their valuable suggestions to revise the original paper.

References

[1] C. Alsina, G. Mayor, M.S. Tomas and J. Torrens, A characterization of a class of aggregation functions, Fuzzy Sets and Systems 53 (1993) 33-38

[2] J. C. Bezdek and T. D. Harris, Fuzzy partitions and relations: An axiomatic basis for clustering, Fuzzy Sets and Systems 1 (1978) 111-127

[3] P. Burillo, M.L. Eraso and M.I. Portilla, Fuzzy n-ary compositions based on aggregation operators, Proc. of the 6th IEEE International Conference on Fuzzy Systems, Barcelona (1997) 1671-1675

[4] P. Burillo, M.L. Eraso and M.I. Portilla, Some properties of the composition based on aggregation operators, Proc. of the 2nd International Fuzzy Based Experts Systems Congress, Sofia (1996) 46-50

[5] C. Chen-Tung and H. Hsi-Mei, Aggregation of fuzzy opinions under group decision making, Fuzzy Sets and Systems 79 (1996) 279-285

[6] W. Cholewa, Aggregation of fuzzy opinions: An axiomatic approach, Fuzzy Sets and Systems 17 (1985) 249-258

[7] A. Di Nola, S. Sessa, W. Pedrycz and E. Sanchez, Fuzzy relation equations and their applications to knowledge engineering (Kluwer Academic Publishers, 1989)

[8] J. Dombi, Basic concepts for a theory of evaluation: The aggregative operator, European Journal Operation Research 10 (1982) 282-293

[9] D. Dubois and J. L. Koning, Social choice axioms for fuzzy set aggregation, Fuzzy Sets and Systems 58 (1991) 339-342

[10] J. Fodor and M. Roubens, Fuzzy preference modelling and multicriteria decision support (Kluwer Academic Publishers, 1994)

[11] L. W. Fung and K. S. Fu, An axiomatic approach to rational decision making in a fuzzy environement, in: L. A. Zadeh, K. S. Fu, K. Tanaka and M. Simura Eds., Fuzzy Sets and their applications to cognitive an decision processes (Academic Press, 1975) 227-256

[12] G. J. Klir and B. Yuan, Fuzzy sets and Fuzzy Logic. Theory and Applications (Prentice-Hall International, 1995)

[13] R. Krishnapuram and L.I. Kuncheva, A fuzzy consensus aggregation operator, Fuzzy Sets and Systems 79 (1996) 347-356

[14] G. Mayor and J. Torrens, On a class of binary operations: non-strict Archimedean aggregation functions, Proc. of the 18th ISMVL, Palma de Mallorca (1988) 54-59

[15] M. Miyakoshi and M. Shimbo, Solutions of composite fuzzy relational equations with triangular norms, Fuzzy Sets and Systems 16 (1985) 53-63

[16] M. Mizumoto, Pictorial Representations of fuzzy connectives, Part I: Cases of t-norms, t-conorms and averating operators, Fuzzy Sets and Systems 31 (1989) 217-242

[17] F. J. Montero, Aggregation of fuzzy opinions in a non-homogeneous group, Fuzzy Sets and Systems 25 (1988) 15-20

[18] M.I. Portilla, P. Burillo and M.L. Eraso, Properties of the fuzzy composition based on aggregation operators, Fuzzy Sets and Systems, to appear (accepted October 1997)

[19] R. Ramakrishnan and C. J. M. Rao, The fuzzy weighted additive rule, Fuzzy Sets and Systems 46 (1992) 177-187

[20] A. Rosenfeld, Fuzzy graphs, in: L. A. Zadeh et al., Fuzzy Sets and their applications to Cognitive and Decision Processes (Academic Press, 1975) 77-95

[21] S. Weber, A general concept of fuzzy connectives, negations and implications based on t-norms and t-conorms, Fuzzy Sets and Systems 11 (1983) 115-134

[22] R. R. Yager, Families of OWA operators, Fuzzy Sets and Systems 59 (1993) 125-148

[23] H.J. Zimmermann and P. Zysno, Decisions and evaluations by Hierarchical aggregation of information, Fuzzy Sets and Systems 10 (1983) 243-260

UNA FAMILIA DE MEDIDAS DIFUSAS ASOCIADA A CADA EVIDENCIA *

José A. Herencia
Dpto. Informática y Análisis Numérico
Campus de Rabanales (Edif. C-2)
Universidad de Córdoba
jaherencia@uco.es

M. Teresa Lamata
Dpto. Ciencias de la Computación e I.A.
E.T.S. Ingeniería Informática
Universidad de Granada
mtl@decsai.ugr.es

Resumen

A partir de cualquier asignación básica de probabilidad definida sobre un conjunto finito, se genera una familia de medidas difusas a la que pertenecen la creencia y la plausibilidad propias de tal evidencia. Tanto la definición como el estudio de ciertas propiedades de dicha familia se obtiene mediante funciones crecientes del intervalo unidad en sí mismo.

Palabras clave: Asignación Básica de Probabilidad, Evidencia, Posibilidad, Probabilidad, Medida Difusa.

1 PRELIMINARES Y NOTACIÓN

En este trabajo consideramos conjuntos finitos $X = \{x_1, x_2, \ldots, x_n\}$. Denotamos por $\mathcal{P}(X)$ al conjunto de partes de X. Para cualquier subconjunto $A \in \mathcal{P}(X)$, consideramos su cardinal $|A|$ y su complementario $A' := X \setminus A$. También usaremos el conjunto $\mathbf{N}$ de números enteros positivos.

Se denomina **medida difusa** sobre X a cualquier aplicación $g : \mathcal{P}(X) \to [0,1]$ que sea creciente ($A \subseteq B \Rightarrow g(A) \leq g(B)$) y que tome los dos valores siguientes: $g(\emptyset) = 0, g(X) = 1$. Dada cualquier medida difusa g sobre X, su **medida dual** es la medida difusa g', definida como sigue: $g'(A) := 1 - g(A')$, $\forall A \subseteq X$. Es inmediato comprobar que $g'' = g$, por lo que toda medida difusa puede estudiarse junto a su dual; denominándose g **autodual** en caso de ser $g' = g$. La mayoría de las medidas difusas que se manejan prácticamente determinan **parejas ordenadas**, en el sentido de que g y g' son comparables: siempre es una de ellas menor o igual que la otra. Para la clasificación de los distintos tipos de medidas difusas nos remitimos a [5] y [6].

Se denomina **asignación básica de probabilidad (A.B.P.)** sobre X a cualquier aplicación $m : \mathcal{P}(X) \to [0,1]$ que verifique las dos condiciones siguientes:

$$m(\emptyset) = 0, \qquad \sum_{A \subseteq X} m(A) = 1.$$

Denotamos por $\mathcal{M}_X$ al conjunto de asignaciones básicas de probabilidad definidas sobre el conjunto X. Dada cualquier $m \in \mathcal{M}_X$, se denominan **elementos focales** de m a los subconjuntos B de X que cumplen la condición $m(B) > 0$. Asociada a tal m se considera la pareja de medidas difusas ordenadas formada por la **creencia** y la **plausibilidad**, definidas respectivamente como sigue ($\forall A \subseteq X$):

$$Bel(A) := \sum_{B \subseteq A} m(B), \qquad Pl(A) := \sum_{B \cap A \neq \emptyset} m(B)$$

A partir de la función Bel, se puede calcular la única A.B.P. m de la que es medida de creencia, mediante la fórmula siguiente ($\forall A \subseteq X$):

$$m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} Bel(B)$$

Por tanto, podemos obtener las tres funciones m, Bel y Pl una vez conocida cualquiera de ellas (ya que Bel y Pl son medidas duales). La información que aporta cualquiera de estas funciones es denominada una **evidencia** sobre X. Hay dos tipos de evidencia especialmente importantes:

(1). Cuando los elementos focales forman una cadena (respecto a la inclusión de conjuntos), tenemos una evidencia **consonante**, que está determinada por la **distribución de posibilidad** $\pi : X \to [0,1]$ siguiente: $\pi(x) := \sum_{x \in B} m(B)$, $\forall x \in X$. En este caso, las funciones Bel y Pl se denominan, respectivamente, medidas de **necesidad** y **posibilidad**, siendo ($\forall A \subseteq X$):

$$Bel(A) = \min_{x \notin A} (1 - \pi(x)), \qquad Pl(A) = \max_{x \in A} \pi(x)$$

La **ignorancia [total]** corresponde al caso particular en que toda la evidencia es asignada al conjunto X (es decir: $m(X) = 1$), no aportando ninguna información respecto a sus subconjuntos estrictos.

* Financiado por la DGICYT, Proyecto PB95-1181

(2). La evidencia es **probabilística** cuando todos los elementos focales de m son unitarios. En este caso $p(x) := m(\{x\})$, $\forall x \in X$, es una **distribución de probabilidad** y se cumple que ($\forall A \subseteq X$):

$$Bel(A) = Pl(A) = \sum_{x \in A} p(x)$$

Se tiene **certeza** en el caso particular de que toda la evidencia esté concentrada en un solo elemento: $\exists x \in X$, $m(\{x\}) = 1$.

Cuando hay algún elemento focal B no unitario, la *"masa de probabilidad"* $m(B)$ se asigna a todo el conjunto B sin que se atribuya una cantidad determinada de la misma a cada uno de sus elementos. Esto hace que la creencia no sea aditiva (al ser para B mayor que la suma de creencias de sus elementos), por lo que no se trata de una probabilidad. No obstante, se considera a veces una probabilidad asociada a cualquier A.B.P., obtenida de la siguiente forma: la masa de probabilidad $m(B)$ de cada elemento focal se reparte uniformemente entre todos sus elementos. Así, en la distribución de probabilidad obtenida, se suma a cada elemento la probabilidad $m(B)/|B|$. En definitiva, resulta la siguiente medida de probabilidad:

$$Pw(A) := \sum_{m(B)>0} m(B)\frac{|B \cap A|}{|B|}, \quad \forall A \subseteq X. \qquad (1)$$

Partiendo de esta expresión, consideramos una generalización de la misma en la Sección 2, resultando una familia de medidas difusas que contiene (como casos extremos) la creencia y la plausibilidad. Dicha familia se obtiene mediante funciones crecientes del intervalo unidad en sí mismo; de forma que algunas propiedades de estas funciones inducen propiedades en las medidas difusas que originan, como vemos en la Sección 3. Ahí también consideramos algunos casos particulares y ejemplos, concluyendo con algunas cuestiones abiertas.

2 DEFINICIÓN DE LA FAMILIA DE MEDIDAS DIFUSAS $\{f_\rho\}_{\rho \in \mathcal{I}_0}$

Al estudiar medidas de entropía asociadas a evidencias [1, 2] aparecen bajo un mismo tratamiento las medidas difusas Bel, Pl y Pw, ya que la *aleatoriedad* de m puede determinarse por la expresión:

$$H(m) = -\sum_{m(B)>0} m(B) \log f(B). \qquad (2)$$

La cual corresponde a:

- La *medida de confusión* de Höhle [3] si $f = Bel$.
- La *medida de disonancia* de Yager [7] si $f = Pl$.
- La *medida de discordancia (o conflicto)* de Klir y Ramer [4] cuando $f = Pw$.

Esto nos ha motivado a explicitar una formulación que permita obtener una familia de medidas difusas que englobe a las tres citadas. Para ello nos basamos en los hechos siguientes. Consideremos cualquier A.B.P. m sobre X y cualquier $A \subset X$. Entonces, los elementos focales B que no cortan a A (para los cuales $|B \cap A| = 0$) se descartan en el cálculo de $Bel(A)$ y de $Pl(A)$. Por el contrario, aparecen en ambos cálculos los elementos focales B que están incluidos en A (para los cuales $|B \cap A| = |B|$). La diferencia entre $Bel(A)$ y $Pl(A)$ se debe a los elementos focales B "intermedios", que cortan a A pero no están incluidos en A (es decir, los elementos focales B para los que $0 < |B \cap A| < |B|$). En la medida de probabilidad Pw, considerada en (1), se asigna el peso $\frac{|B \cap A|}{|B|}$ a cada elemento focal B. De aquí surge la generalización consistente en transformar el peso $\frac{|B \cap A|}{|B|}$, cambiándolo por el factor $\rho\left(\frac{|B \cap A|}{|B|}\right)$, siendo $\rho : [0,1] \to [0,1]$ cualquier función creciente. Concretando, tenemos:

DEFINICIÓN 2.1 *Denominemos $\mathcal{I}$ a la familia de funciones crecientes de $[0,1]$ en $[0,1]$. Para cualquier A.B.P. $m \in \mathcal{M}_X$ y cualquier $\rho \in \mathcal{I}$, consideramos la función $f_\rho : \mathcal{P}(X) \to [0,1]$ definida por:*

$$f_\rho(A) := \sum_{B \neq \emptyset} m(B)\rho\left(\frac{|B \cap A|}{|B|}\right), \quad \forall A \subseteq X. \qquad (3)$$

PROPOSICIÓN 2.2 *Para cualquier A.B.P. $m \in \mathcal{M}_X$ y cualquier $\rho \in \mathcal{I}$, se cumple:*

1. *Las condiciones necesarias y suficientes para que f_ρ sea una medida difusa sobre X son las siguientes: $\rho(0) = 0$ y $\rho(1) = 1$. Por eso consideramos el conjunto $\mathcal{I}_0 := \{\rho \in \mathcal{I} : \rho(0) = 0,\ \rho(1) = 1\}$.*

2. *Si $\rho \in \mathcal{I}_0$, entonces la medida dual de f_ρ es:*

   $f'_\rho = f_{\rho'}$ donde $\rho'(x) = 1 - \rho(1-x)$, $\forall x \in [0,1]$.

DEMOSTRACIÓN:

1. Si en la ecuación (3) tomamos $A = \emptyset$ (resp. $A = X$) se obtiene la suma, extendida a todos los elementos focales B, de $\rho(0)m(B)$ (resp. $\rho(1)m(B)$). Por tanto: $f_\rho(\emptyset) = \rho(0)$ y $f_\rho(X) = \rho(1)$. Teniendo en cuenta estas igualdades, sólo falta comprobar que el crecimiento de ρ implica el crecimiento de f_ρ. En efecto, consideremos subconjuntos cualesquiera tales que $A_1 \subseteq A_2 \subseteq X$. Entonces, para cualquier $B \neq \emptyset$ se cumple que $\frac{|B \cap A_1|}{|B|} \leq \frac{|B \cap A_2|}{|B|}$. De donde, por ser ρ creciente, resulta efectivamente que $f_\rho(A_1) \leq f_\rho(A_2)$.
2. Partiendo de que f_ρ sea una medida difusa, basta

calcular su dual: $f'_\rho(A) = 1 - f_\rho(A') =$

$$= \sum_{B\neq\emptyset} m(B) - \sum_{B\neq\emptyset} m(B)\rho\left(\frac{|B\cap A'|}{|B|}\right).$$

Usando la igualdad $|B\cap A'| = |B| - |B\cap A|$ y asociando términos obtenemos: $f'_\rho(A) =$

$$= \sum_{B\neq\emptyset} m(B)\left[1-\rho\left(1-\frac{|B\cap A|}{|B|}\right)\right] = f_{\rho'}(A). \ \square$$

COROLARIO 2.3 *Asociadas con cualquier A.B.P. $m \in \mathcal{M}_X$, podemos considerar las siguientes medidas difusas:*

1. *Para cada $\alpha \in [0,1]$, tenemos la medida difusa P_α definida por ($\forall A \subseteq X$):*

$$P_\alpha(A) := \sum_{B\subseteq A} m(B) + \alpha \sum_{\substack{B\not\subseteq A \\ B\cap A\neq\emptyset}} m(B) \qquad (4)$$

Su medida dual es: $P'_\alpha = P_{1-\alpha}$. En particular, $P_0 = Bel$ y $P_1 = Pl$.

2. *Para cada $k \in \{2i-1 : i \in \mathbf{N}\} \cup \{\frac{1}{2i-1} : i \in \mathbf{N}\}$ tenemos la medida difusa autodual $P_{[k]}$ definida por ($\forall A \subseteq X$):*

$$P_{[k]}(A) := \sum_{B\neq\emptyset} \frac{m(B)}{2}\left[1+\left(2\frac{|A\cap B|}{|B|}-1\right)^k\right]$$

En particular, tenemos la probabilidad $Pw = P_{[1]}$ considerada en (1).

DEMOSTRACIÓN:

1. Basta considerar la medida $P_\alpha = f_{\rho_\alpha}$, donde $\rho_\alpha \in \mathcal{I}_0$ está definida por:

$$\rho_\alpha(x) := \begin{cases} 0, & x = 0 \\ \alpha, & 0 < x < 1 \\ 1, & x = 1 \end{cases} \qquad (5)$$

Además es inmediato comprobar que $\rho'_\alpha = \rho_{1-\alpha}$, $P_0 = Bel$ y $P_1 = Pl$.

2. En este caso se trata de la medida $P_{[k]} = f_{\rho_{[k]}}$, siendo $\rho_{[k]}(x) = \dfrac{1+(2x-1)^k}{2}$. También se comprueba fácilmente que $\rho_{[k]} \in \mathcal{I}_0$ y que $\rho'_{[k]} = \rho_{[k]}$. En el caso particular $k = 1$, se tiene simplemente que $\rho_{[1]} = Id$ es la función identidad en $[0,1]$, que origina la probabilidad Pw. $\square$

NOTA: Es obvio que podemos considerar muchas otras familias de medidas difusas, sin más que tomar las funciones $\rho \in \mathcal{I}_0$ adecuadas. Algunos ejemplos son:

- Para cada $r \in (0,+\infty)$ podemos tomar la función $\rho(x) = x^r$, obteniendo la siguiente medida difusa:

$$P_{(r)}(A) := \sum_{B\neq\emptyset} m(B)\left(\frac{|A\cap B|}{|B|}\right)^r$$

De nuevo tenemos el caso particular $Pw = P_{(1)}$.

- Para cada $b \in (1,+\infty)$ tenemos la medida difusa correspondiente a la función $\rho(x) = \dfrac{b^x-1}{b-1}$.
- También con $b \in (1,+\infty)$ podemos tomar $\rho(x) = \log_b[1+(b-1)x]$.

Al contrario de lo que ocurría en las familias de funciones consideradas en el corolario 2.3, estas nuevas familias de funciones no contienen a sus respectivas duales. Observemos también que siempre que $\rho \in \mathcal{I}_0$ sea estrictamente creciente, podemos considerar la función $\rho^{-1} \in \mathcal{I}_0$ (siendo éste el procedimiento que relaciona los dos últimos ejemplos). $\triangleright$

3 ALGUNAS PROPIEDADES DE LAS MEDIDAS DIFUSAS f_ρ

Vemos ahora cómo algunas propiedades de las funciones $\rho \in \mathcal{I}_0$ se transfieren a las correspondientes medidas difusas.

PROPOSICIÓN 3.1 *Para cualesquiera $\rho, \hat\rho \in \mathcal{I}_0$ se cumple:*

1. *$\forall\beta \in [0,1]$, $\beta\rho + (1-\beta)\hat\rho \in \mathcal{I}_0$, siendo $f_{\beta\rho+(1-\beta)\hat\rho} = \beta f_\rho + (1-\beta) f_{\hat\rho}$.*
2. *Si $\rho \leq \hat\rho$ entonces $f_\rho \leq f_{\hat\rho}$.*
3. *Si la gráfica de ρ es simétrica respecto al punto $\left(\frac{1}{2},\frac{1}{2}\right)$, entonces f_ρ es autodual.*

DEMOSTRACIÓN: Los dos primeros apartados se deducen inmediatamente de la ecuación (3). Para demostrar el tercero, al ser $f'_\rho = f_{\rho'}$ con $\rho'(x) = 1-\rho(1-x)$ (Proposición 2.2, apartado 2), basta comprobar que la simetría enunciada implica la igualdad $\rho' = \rho$. En efecto, tal simetría consiste en que $\forall u \in [-1/2, 1/2]$, $\rho(1/2+u) - 1/2 = 1/2 - \rho(1/2-u)$, lo que equivale a que $\forall u \in [-1/2,1/2]$, $1-\rho(1/2+u) = \rho(1/2-u)$. Entonces basta efectuar el cambio de variable $x = 1/2 - u$ para obtener la condición buscada: $\forall x \in [0,1]$, $1-\rho(1-x) = \rho(x)$. $\square$

Estos resultados indican la posibilidad de obtener nuevas funciones $\rho \in \mathcal{I}_0$, así como las correspondientes medidas difusas f_ρ, mediante combinaciones lineales convexas. Por ejemplo, podemos considerar las siguientes medidas ($\forall\alpha,\beta \in [0,1]$):

$$\alpha Pl + (1-\alpha)Bel = P_\alpha, \quad \beta Pw + (1-\beta)P_\alpha, \quad etc.$$

Asimismo, los resultados anteriores corroboran hechos conocidos como la igualdad $Pw' = Pw$ o las desigualdades siguientes:

$$m \leq Bel \leq Pw \leq Pl, \qquad Bel \leq f_\rho \leq Pl, \ \forall \rho \in \mathcal{I}_0.$$

De la última desigualdad se deduce el hecho obvio siguiente: cuando $m \in \mathcal{M}_X$ es una evidencia probabilística, entonces la familia $\{f_\rho\}_{\rho \in \mathcal{I}_0}$ se reduce a la medida de probabilidad determinada por m. Esto motiva la cuestión de buscar condiciones sobre m y sobre ρ que garanticen la obtención de determinados tipos de medidas f_ρ. En este sentido, tenemos:

COROLARIO 3.2 *Para cualquier $\rho \in \mathcal{I}_0$ se cumple:*

1. *Si $\rho(x)+\rho(1-x) \leq 1$, $\forall x \in [0,1]$, entonces resulta la pareja ordenada $f_\rho \leq f'_\rho$.*
2. *Si $\rho(x)+\rho(1-x) \geq 1$, $\forall x \in [0,1]$, entonces resulta la pareja ordenada $f_\rho \geq f'_\rho$.*

DEMOSTRACIÓN: Basta aplicar el segundo apartado de la Proposición 3.1 ya que la desigualdad $\rho(x) + \rho(1-x) \leq 1$ (resp. $\rho(x) + \rho(1-x) \geq 1$) implica que $\rho \leq \rho'$ (resp. $\rho \geq \rho'$). □

Para ilustrar lo anterior podemos considerar cualquier $\alpha \in [0, 1/2]$, en cuyo caso la definición (5) muestra claramente que la función ρ_α cumple las condiciones del primer apartado, resultando efectivamente la pareja ordenada de medidas difusas siguiente: $P_\alpha \leq P_{1-\alpha}$. Vemos ahora un ejemplo donde se muestra que si ρ y ρ' no están ordenadas, entonces f_ρ y f'_ρ pueden no estarlo tampoco; lo que impide que f_ρ sea una medida que junto con su dual acoten una probabilidad (no pudiendo por tanto ser representable ni mucho menos una evidencia):

EJEMPLO: Sea X un conjunto con cardinal $n = 4$, donde definimos la evidencia consonante dada por la siguiente A.B.P: $m(\{x_1\}) = m(\{x_1, x_2\}) = 1/4$, $m(X) = 1/2$. Tomamos además la función ρ definida como sigue:

$$\rho(x) := \begin{cases} 0, & x = 0 \\ 1/4, & 0 < x \leq 1/2 \\ 1, & 1/2 < x \leq 1 \end{cases}$$

Observemos que ρ verifica la condición del segundo apartado del Corolario 3.2 en todos los puntos excepto en $x = 1/2$. Pero eso es suficiente para que f_ρ y f'_ρ no estén ordenadas, ya que basta calcular los siguientes valores: $f_\rho(\{x_3\}) = 1/8 > 0 = f'_\rho(\{x_3\})$, mientras que $f_\rho(\{x_1, x_2\}) = 5/8 < 7/8 = f'_\rho(\{x_1, x_2\})$. ▷

Finalmente consideramos otro ejemplo para mostrar que ni la probabilidad Pw ni la medida difusa $P_{1/2}$ determinan unívocamente la A.B.P m de la cual provienen (hecho que sí ocurre con Bel y Pl como es sabido y con P_α, si $\alpha \neq 1/2$, como puede comprobarse).

EJEMPLO: Sea $n = 2$, $r \in [0, 1/2]$ y consideremos la A.B.P. definida por: $m(\{x_1\}) = m(\{x_2\}) = r$, $m(X) = 1 - 2r$. Entonces basta efectuar simples cálculos para comprobar que: $\forall r \in [0, 1/2]$, $Pw(\{x_1\}) = Pw(\{x_2\}) = 1/2 = P_{1/2}(\{x_1\}) = P_{1/2}(\{x_2\})$. ▷

Las propiedades y ejemplos anteriores sugieren nuevas cuestiones (que indican una posible continuación de este trabajo). Por ejemplo:

- ¿Es inyectiva la aplicación que asigna a cada ρ la medida difusa f_ρ?
- Además de Bel y Pl, ¿cuáles son todas las medidas de la familia $\{f_\rho\}_{\rho \in \mathcal{I}_0}$ que permiten reconstruir la A.B.P. m de la cual provienen? ¿A qué tipo de funciones ρ corresponden tales medidas?

NOTA: Una vez fijada la A.B.P. m, en la determinación de f_ρ intervienen solamente los valores $\rho\left(\frac{|B \cap A|}{|B|}\right)$ correspondientes a los distintos elementos focales B. Por tanto, al responder las cuestiones anteriores, deberemos considerar una relación de equivalencia en $\mathcal{I}_0$ de forma que cada ρ venga determinada exclusivamente por los valores que toma en el conjunto de puntos $\{\frac{p}{|B|} : 0 \leq p \leq |B|, B$ es elemento focal de $m\}$ (agradecemos a un referee la determinación de este conjunto). ▷

Referencias

[1] J.A. Herencia y M.T. Lamata, Una generalización de la entropía a través de la Teoría de Dempster–Shafer, *VI Congreso Español Sobre Tecnologías y Lógica Fuzzy* (Oviedo, 1996). Actas, pp. 143-147.

[2] J.A. Herencia and M.T. Lamata, Entropy Measures Associated with a Fuzzy Basic Probability Assignment, *Proc. of the FUZZ-IEEE'97* (I.I.I.A., Barcelona, Spain, 1997) 863-868.

[3] U. Höhle, Fuzzy plausibility measures. In: E.P. Klement (ed.), *Proc. Third Intern. Seminar on Fuzzy Set Theory* (Johannes Kepler Univ., Linz, 1981) 7-30.

[4] G.J. Klir and A. Ramer, Uncertainty in Dempster-Shafer theory: A critical re-examination, *Int. J. Gen. Syst.* 18 (1990) 155-166.

[5] M. Lamata and S. Moral, Classification of fuzzy measures, *Fuzzy Sets and Systems* 33 (1989) 243-253.

[6] J.F. Verdegay-López, Tratamiento de la incertidumbre mediante medidas difusas. En *Algunos aspectos del tratamiento de la Información en Inteligencia Artificial* (Publicaciones de la Universidad de Granada, 1991) 155-172.

[7] R.R. Yager, Entropy and specificity in a Mathematical Theory of Evidence, *Int. J. Gen. Syst.* 9 (1983) 249-260.

Representación de conceptos difusos mediante índices semánticos.

L.J.Linares
Dep. Informática. ESI . UCLM.
13071 Ciudad Real. España
ljimenez@inf-cr.uclm.es

J.A.Olivas
Dep. Informática. ESI . UCLM.
13071 Ciudad Real. España
jaolivas@inf-cr.uclm.es

Resumen

Magnitudes inciertas o imprecisas son representadas habitualmente mediante conjuntos difusos. Estos conjuntos se definen por medio de una función de pertenencia $\mu : X \rightarrow [0,1]$ sobre los elementos de un referencial X. Esta representación funcional de los conceptos no nítidos obliga al uso del principio de extensión para transladar operaciones o conceptos definidos en otros dominios. El uso del principio de extensión basado en el concepto de alfa-corte de un conjunto difuso origina en algunos casos la necesidad de recurrir a difíciles cálculos funcionales. En este trabajo se presenta una representación alternativa y compatible con la teoría de conjuntos difusos, definiendo un conjunto de características semánticas propias del carácter impreciso de este tipo de conceptos, que tratadas conjuntamente permiten manipular esos conceptos de una forma alternativa al principio de extensión.

Palabras clave: Índices semánticos, prototipos.

1 Índices asociados a conjuntos difusos.

Varios son los autores que han descrito una serie de índices reales, con los que reflejar la información representada por un conjunto difuso. Ejemplo de estos índices, son los criterios de desborrosificación del máximo, la media del máximo o el centro de masas $CM(\mu)$ de un conjunto difuso (establecido sobre un referencial X) , definido por

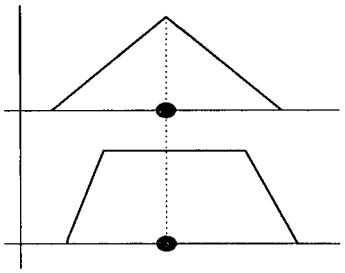


Figura 1: Centro de masas de distintos conjuntos difusos

$$CM(\mu) = \frac{\int \mu(x)xdx}{\int \mu(x)dx} \quad (1)$$

reflejando un único valor real para la totalidad del conjunto difuso.

Este tipo de simplificaciones son efectivas para algunos tipos de aplicaciones como son el control difuso y el ranking de conjuntos difusos; sin embargo, no son válidas para otros problemas que necesitan más información. Un problema esencial en este tipo de índices únicos, es la imposibilidad de extraer de ellos el conjunto difuso que los originó. En la figura 1 se representan dos conjuntos difusos con características sustancialmente distintas que poseen el mismo valor para cualquiera de los índices señalados.

Otro de los problemas que posee el índice del CM, es que puede desplazar el grueso de la información aportada por el conjunto difuso hacia valores anormales o donde el grado de pertenencia de los elementos del dominio no es el más representativo. Un ejemplo se puede observar en la figura 2 cuando

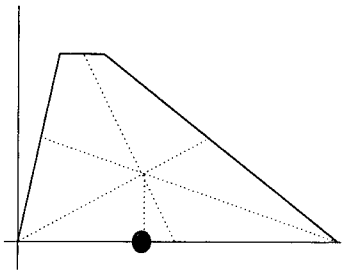


Figura 2: Centro de masas con valor de pertenencia inferior a 1.

el $CM(\mu)$ se sitúa fuera de donde se encuentran los valores de máxima pertenencia .

Este efecto se vería amplificado si al conjunto difuso se le asociase una etiqueta, y a ésta como valor representativo o nominal su CM.

Mediante otro enfoque M. Delgado, M. A. Vila y W. Voxman [3][4] proponen una representación canónica para los números difusos como un par de índices reales $(V_s(\mu), A_s(\mu))$, (Valor, Ambigüedad), definidos como :

$$V_s(\mu) = \int_0^1 s(r)\,[L_\mu(r) + R_\mu(r)]\,dr \qquad (2)$$

$$A_s(\mu) = \int_0^1 s(r)\,[R_\mu(r) - L_\mu(r)]\,dr \qquad (3)$$

donde s(r) es una función de reducción, definida como una función creciente $s : [0,1] \rightarrow [0,1]$ con las propiedades s(0)=0 y s(1)=1, que permite ponderar la influencia de los r-cortes que definen a $L_\mu(r)$ y $R_\mu(r)$ como:

$$L_\mu(r) = \begin{cases} inf\,\{x | x \in \mu_r\} & x \in (0,1] \\ inf\,\{x | x \in Supp(\mu)\} & r = 0 \end{cases} \qquad (4)$$

$$R_\mu(r) = \begin{cases} sup\,\{x | x \in \mu_r\} & x \in (0,1] \\ sup\,\{x | x \in Supp(\mu)\} & r = 0 \end{cases} \qquad (5)$$

siendo $\mu_r = \{x | \mu(x) \geq r\}$ y $Supp(\mu) = \{x | \mu(x) > 0\}$.

El valor de un número difuso $V_s(\mu)$ representa el valor central de una magnitud definida mediante el número difuso, mientras que la ambigüedad $A_s(\mu)$ representaría "la forma" del mismo.

Esta representación, mediante el par de índices anteriormente descritos, tiene la ventaja que determinados números difusos, aquellos que están centrados, pueden reconstruirse a partir de sus índices. La generalización de este reconstrucción inversa para cualquier número difuso obliga a la sustitución del índice de ambigüedad $A_s(\mu)$ por otros dos índices $A_L(\mu)$ y $A_R(\mu)$, definidos en función de $q = (L_\mu(1) + R_\mu(1))/2$, punto central del intervalo modal del conjunto difuso μ, como

$$A_L(\mu) = \int_0^1 s(r)\,[q - L_\mu(r)]\,dr \qquad (6)$$

$$A_R(\mu) = \int_0^1 s(r)\,[R_\mu(r) - q]\,dr \qquad (7)$$

2 Representación de un número difuso mediante sus índices semánticos.

Desde nuestro punto de vista, son varios los conceptos que quedan reflejados mediante un conjunto difuso. Centrándose en las magnitudes expresadas por medio de un número difuso trapezoidal, se pueden destacar los siguientes:

1. *Generalidad de la magnitud.* Una magnitud representada mediante un conjunto difuso será más o menos general conforme sea mayor o menor el número de elementos del referencial sobre los que dicha magnitud es relevante. El valor que define dicha generalidad $G(\mu)$ será $G(\mu) = (d - a)$ siendo $a = inf\{Supp(\mu)\}$ y $d = sup\{Supp(\mu)\}$.

2. *Imprecisión de una magnitud.* Una magnitud representada mediante un conjunto difuso será más o menos imprecisa conforme sea mayor o menor el número de elementos del referencial sobre los que dicha magnitud está perfectamente definida. El valor que define dicha imprecisión será $I(\mu) = (c - b)$ siendo $b = inf\{Mod(\mu)\}$, $c = sup\{Mod(\mu)\}$ y $Mod(\mu) = \{x | \mu(x) = 1\}$.

3. *Ambigüedad de una magnitud.* Relación entre la imprecisión y generalidad de la magni-

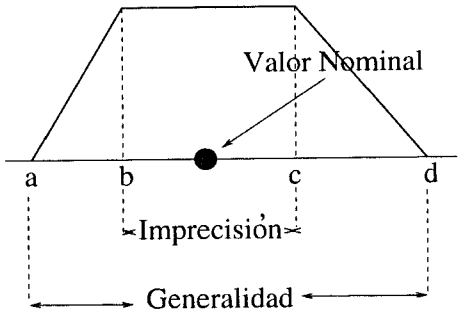


Figura 3: Índices semáticos

tud expresada como $A(\mu) = I(\mu)/G(\mu)$, ó 0 si $G(\mu) = 0$.

4. *Valor nominal de una magnitud.* Valor del conjunto $Mod(\mu)$ elegido como representante de la magnitud en toda su perfección, se definirá como $VN(\mu) = (c - b)/2$.

En la figura 3 se muestran estas características para el caso de un número difuso trapezoidal.

Se define la *representación semántica* de un número difuso como:

$$RS([a,d], VN, A) \tag{8}$$

El intervalo [a,d] proporciona la información concisa sobre la generalidad de la magnitud, que junto a los valores VN y A establecen el resto de valores destacable como el de la imprecisión I=A*(d-a).

Dado un número trapezoidal T(a,b,c,d) su representación semántica es única y viene definida como

$$RS\left([a,d], \frac{b+c}{2}, \frac{c-b}{d-a}\right) \tag{9}$$

Esta representación es reversible, en el sentido de que a partir de un representación semántica dada $RS([a,d], VN, A)$ se puede obtener el número difuso trapezoidal resolviendo el siguiente par de ecuaciones

$$\begin{aligned} b + c &= 2VN \\ c - d &= A(d-a) \end{aligned} \tag{10}$$

Operando en (10) se tiene que

$$\begin{aligned} c &= VN + \tfrac{A(d-a)}{2} \\ b &= VN - \tfrac{A(d-a)}{2} \end{aligned} \tag{11}$$

Resultando el número trapezoidal

$$T(a, VN - \frac{A(d-a)}{2}, VN + \frac{A(d-a)}{2}, d)$$

cuya función de pertenecia $\mu(x)$ se define como

$$\mu(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{VN - \frac{A(d-a)}{2} - a} & a \leq x < VN - \frac{A(d-a)}{2} \\ 1 & VN - \frac{A(d-a)}{2} \leq x \\ & x \leq VN + \frac{A(d-a)}{2} \\ \frac{d-x}{d - \left(VN + \frac{A(d-a)}{2}\right)} & VN + \frac{A(d-a)}{2} < x \leq d \\ 0 & d < x \end{cases} \tag{12}$$

3 Prototipo semántico de una colección de valores no nítidos.

Cuando se realiza un análisis de algún tipo de información, parece de especial relevancia la capacidad de determinar un objeto típico o prototipo de una colección de objetos. De igual forma, la idea de la tipicalidad de un valor o valor típico juega un papel importante para el razonamiento por sentido común, es más, estos valores son muchas veces utilizados como valores por defecto.

Es habitual la necesidad de determinar un valor típico en situaciones en las que se dispone de una colección de observaciones que queremos categorizar o extraer las características que las identifican. Yager[5] expresa el concepto de prototipo o valor típico como " *el valor que es él mismo o muy similar para muchas de las observaciones que estamos intentando tipificar.*"

Sea el concepto de prototipo de una colección de elementos, aquel elemento que posee unas características que reflejan a la generalidad de la colección.

De manera más formal, considérese una colección $L = \{e_1, \ldots, e_m\}$ de m elementos, cada uno

de ellos $e_i = (x_1^i, \ldots, x_n^i)$ definidos mediante n características $X_1 \times \ldots \times X_n$. Sea d_i una medida de distancia de los valores de la característica i-esima. Se definirá el prototipo $Prot(L) = (p_1, \ldots, p_n)$ como el conjunto de valores, de las características que minimizan las distancias entre los valores posibles de cada característica:

$$d_i(x, p_i) \leq d(x, y) \, \forall x, y \in X_i \tag{13}$$

Si se toma una colección de elementos no nítidos, definidos mediante una representación semántica

$$e_i = RS_i([a_i, d_i], VN_i, A_i) \tag{14}$$

Se define el prototipo Prot(L), ecuación (15), como

$$RS\left(\left[\frac{1}{m}\sum_{i=1}^{m} a_i, \frac{1}{m}\sum_{i=1}^{m} d_i\right], \frac{1}{m}\sum_{i=1}^{m} VN_i, \frac{1}{m}\sum_{i=1}^{m} A_i\right) \tag{15}$$

donde cada índice semántico es la media de los valores de los índices semánticos de la colección de elementos.

Este enfoque se puede utilizar para determinar el prototipo de una colección de número difusos.

En primer lugar se obtendrá la representación semántica de cada número difuso, luego se calculará el prototipo mediante la ecuación (15); para finalmente reconstruir el número difuso prototipo de la colección.

4 Conclusiones y trabajos futuros

Lo que se ha pretendido es definir y mostrar la utilidad de índices semánticos para la caracterización reversible de conjuntos borrosos representados por números borrosos trapezoidales.

En base a esta definición, se introduce la definición de prototipo de un conjunto de números borrosos, que se intuye más operativa y significativa que las que se usan habitualmente.

La proyección de este trabajo pasa por extender el concepto de índice semántico a otras caracterizaciones de conjuntos difusos, así como hacia una operatividad para manipular sistemas de reglas y sistemas de control con esta concepción.

Finalmente se comprobará el funcionamiento en sistemas reales, aplicaciones en las que ya se han usado técnicas clásicas, tanto de manipulación de conceptos borrosos como de uso de la noción de prototipo derivada de estos conceptos.

Referencias

[1] Kóczy, L. T., Hirota, K.: "Ordering, distance and closeness of fuzzy sets", Fuzzy Sets and Systems 59 (1993) 281-293.

[2] Kóczy, L. T., Hirota, K.: "Approximate Reasoning by Linear Rule Interpolation and General Approximation", International Journal of Approximate Reasoning (1993); 9:197-225.

[3] M Delgado, M. A. Vila, W. Voxman "On canonical representation of fuzzy numbers." Fuzzy Sets and Systems 93 (1998) 125-135.

[4] M Delgado, M. A. Vila, W. Voxman "A fuzzy measure for fuzzy numbers: Applications." Fuzzy Sets and Systems 94 (1998) 205-216.

[5] Ronald R. Yager. "A note on fuzzy measure of typicality." International Journal of intelligent systems. VOL 12, (1997) 233-249

[6] Zhu, Q., Lee, E. S.: "Comparison and Ranking of Fuzzy Numbers", en Fuzzy Regression Analysys, J. Kacprzyk y M. Fedrizzi (eds.), (1992), Physica-Verlag.

UNA ESTRUCTURA ORDENADA EN OPERADORES DE LA MORFOLOGÍA MATEMATICA

P. Burillo
Dpto de Automática y Computación. Universidad Pública de Navarra Campus de Arrosadía 31006. Pamplona. Navarra.
E-mail: pburillo @upna.es

N. Frago
Dpto de Matemática e Informática. Universidad Pública de Navarra Campus de Arrosadía 31006. Pamplona. Navarra.
E-mail: noe.frago @upna.es,

R. Fuentes
Dpto de Automática y Computación. Universidad Pública de Navarra Campus de Arrosadía 31006. Pamplona. Navarra.
E-mail: rfuentes@upna.es

Resumen

En esta comunicación se introducen estructuras de orden en familias de operadores básicos de la Morfología Matemática borrosa y se estudia su estructura.

Palabras claves: Grado de Inclusión, Operador Implicación, Erosión, Morfología Matemática borrosa.

INTRODUCCIÓN

En la Morfología Matemática ordinaria, se distinguen dos tipos de imágenes: las binarias, representadas por subconjuntos *A* del plano, y las imágenes en tonos de gris, representadas como funciones *f* definidas en el plano y con valores en *R*. En ambos casos, el tratamiento para la transformación de las imágenes se modeliza utilizando otras fijas, (los elementos estructurantes), ciertos operadores en el conjunto de imágenes y la composición de ellos. Aparecen como operadores básicos de este tipo: la erosión, dilatación, apertura y cierre. Se demuestra que basta considerar uno de los dos primeros para obtener todos los demás utilizando operaciones conjuntistas en el caso de imágenes binarias y operaciones reticulares en el caso de imágenes con tonos de gris.

La Morfología Matemática Borrosa (MMB) extiende y comprende a la ordinaria al considerar subconjuntos borrosos del plano, (que pueden interpretarse tanto como imágenes vagas o como imágenes en tonos de gris), en lugar de subconjuntos ordinarios o funciones *f* con valores en *R*. Como extensión coherente de la ordinaria, el tratamiento de los subconjuntos borrosos en MMB se realiza también por medio de operadores entre ellos,. Aunque en este caso, en lugar de un único operador erosión, dilatación, etc., aparecen en la literatura distintos operadores erosión borrosa, dilatación borrosa, etc.

Con el propósito de construir operadores erosión borrosa de forma similar a la erosión de la Morfología Matemática binaria, Divyendu y Dougherty [1-4] definen los operadores grado de inclusión para subconjuntos borrosos como relaciones difusas, aplicaciones de F(U) x F(U) en [0, 1], siendo F(U) el conjunto de todos los subconjuntos borrosos de U, verificando condiciones que, para dos subconjuntos difusos A y B de F(U) y una relación difusa R: F(U) x F(U) → [0, 1], permitan considerar a R(A, B) como una medida del grado de inclusión de A respecto de B.

En esta línea, Divyendu y Dougherty [1-4] proponen una colección de axiomas para definir el grado de inclusión. Frago [6] prueba que dichos axiomas no son independientes, dándose la siguiente definición que modifica la propuesta de los autores citados.

Definición 1: Diremos que R: F(U) x F(U) → [0, 1], es un grado de inclusión entre subconjuntos difusos si satisface el siguiente conjunto de axiomas:

IG. 1.- $R(A, B) = 1 \Leftrightarrow A \subseteq B$

IG. 2.- $R(A, B)=0 \Leftrightarrow \exists \{x_n\}_{n\in N} \subseteq U$ tal que $\{A(x_n)\}_{n\in N}$ y $\{1-B(x_n)\}_{n\in N}$ convergen a 1

IG. 3.- R es no decreciente en la segunda variable ($B \subseteq C \Rightarrow R(A, B) \leq R(A, C)$)

IG. 4.- R es no creciente en la primera variable ($B \subseteq C \Rightarrow R(C, A) \leq R(B, A)$)

IG. 5.- $R(A, B) = R(B^C, A^C)$

IG. 6.- $R(A \cup B, C) \geq \min(R(A, C), R(B, C))$

IG. 7.- $R(A, B \cap C) \geq \min(R(A, B), R(A, C))$

En [6] se prueba que para $\lambda : [0, 1] \rightarrow [0, 1]$ la relación borrosa definida para cada $(A, B) \in F(U) \times F(U)$ por

$$R(A, B) = \inf_{x\in U} \{\min(1, \lambda(A(x)) + \lambda(1-B(x)))\},$$

satisface las propiedades de la Definición 1 si y sólo si λ verifica las siguientes condiciones:

Tabla 1

1. - $\lambda(1) = 0$ y $\lambda(0) = 1$,
2.- $p \leq q \Leftrightarrow \lambda(p)+\lambda(1-q) \geq 1$,
3.- λ es decreciente,
4.- λ es continua a la derecha en 0 y a la izquierda en1.

Resulta inmediato comprobar que de las condiciones de la tabla 1, λ resulta ser inyectiva y por lo tanto es estrictamente decreciente.

En [6] se prueba que si g: [0, 0.5] ⟶ [0.5, 1] es una función estrictamente decreciente, continua a trozos y continua en el punto 0 y tal que g(0)=1, entonces la función λ asociada a g definida por:

$$\lambda(p) = \begin{cases} g(p) & \text{si} \quad p \in \left[0, \frac{1}{2}\right] \\ 1 - g(1-p) & \text{si} \quad p \in \left(\frac{1}{2}, 1\right] \end{cases}$$

satisface las cuatro condiciones exigidas en la Tabla 1.

Las imágenes que se consideran son subconjuntos borrosos de un universo **U**, (funciones de **U** en [0, 1]) donde $U = R^2$ o $U = Z^2$ o un subconjunto de estos, (U será Z^2 o un subconjunto de Z^2 en el caso de imágenes digitales). Los elementos de la imagen están caracterizados por las coordenadas espaciales x e y, y un valor numérico comprendido entre el cero y el uno, que nos indica el nivel de gris de la imagen en el punto (x, y). En el caso de imágenes digitales se representan mediante una matriz cuyos índices de filas *i* y columnas *j* representarán las coordenadas espaciales de un punto *(i, j)* de la imagen y el correspondiente valor de dicho elemento de la matriz (número comprendido entre cero y uno) indica el color en dicho punto. Los elementos de la matriz son llamados elementos de la imagen o pixeles.

Con esto, se pueden definir las operaciones de erosión, dilatación, apertura y cierre. Para ello es conocido que si A es un subconjunto borroso, se denota por –A (opuesto de A) al subconjunto borroso definido por $(-A)(x) = A(-x)$ $\forall x \in U$ y se denota por A_z (trasladado de A por $z \in U$) al subconjunto borroso definido por $A_z(x) = A(x-z)$ $\forall x \in U$

Definición 2: [1-4] La erosión (dilatación) de una imagen A por otra B, ambos subconjuntos borrosos pertenecientes a F(U), es otro subconjunto borroso denotado por $\xi(A, B)$ ($\not\otimes(A, B)(z)$), y definido por

$$\xi(A, B)(z) = R(B_z, A)$$
$$(\not\otimes(A, B)(z) = 1 - R((-B)_z, A^c)) \quad \forall z \in \mathbf{U}.$$

Definición 3: [1-4] La apertura (cierre) de un subconjunto borroso A perteneciente a F(U) por otro B, es un subconjunto borroso denotado por @(A, B) ($\zeta(A, B)$), y definido por

$$@(A, B) = \not\otimes(\xi(A, B), B).$$
$$(\zeta(A, B) = \xi(\not\otimes(A, -B), -B)).$$

UN ORDEN PARCIAL $\trianglelefteq$ EN LA CLASE E DE EROSIONES ASOCIADAS A LAS FUNCIONES λ

Consideramos la familia de operadores erosión $E = (\xi_\lambda)$ siendo λ una función que satisface las condiciones de la Tabla 1. Esas erosiones son operadores

$$\xi_\lambda : F(U) \times F(U) \longrightarrow F(U)$$
$$(A, B) \longrightarrow \xi_\lambda(A, B)$$

definidos por: $\forall x \in \mathbf{U}$ $\xi_\lambda(A, B)(z) = R(B_z, A) =$

$$\inf_{x \in \mathbf{U}} \{ \min(1, \lambda(B(x-z)) + \lambda(1 - A(x))) \}$$

o por la expresión equivalente,

$$\xi_\lambda(A, B)(z) = \inf_{x \in \mathbf{U}} \{ \min(1, \lambda(B(x)) + \lambda(1 - A(x+z))) \}.$$

Dicha familia queda descrita de modo unívoco por las funciones g mencionadas en la introducción

Consideramos en E el orden parcial $\trianglelefteq$ dado por

$$\xi_{\lambda_1} \trianglelefteq \xi_{\lambda_2} \Leftrightarrow g_1(x) \le g_2(x) \quad \forall x \in [0, 0.5],$$

en donde g_1 y g_2 generan respectivamente λ_1 y λ_2

Nótese que por las condiciones de construcción de las funciones λ a partir de las funciones g, la condición $g_1(x) \le g_2(x)$ $\forall x \in [0, 0.5]$es equivalente a la verificación de

$$\lambda_1(x) \le \lambda_2(x) \quad \forall x \in [0, 0.5]$$

y

$$\lambda_1(x) \ge \lambda_2(x) \quad \forall x \in (0.5, 1].$$

Los siguientes resultados reflejan algún comportamiento de la aplicación de las erosiones a imágenes.

Teorema 1: Sean dos erosiones borrosas ξ_{λ_1} y $\xi_{\lambda_2} \in E$, tales que $\xi_{\lambda_1} \trianglelefteq \xi_{\lambda_2}$ y A, $B \in F(U)$ dos imágenes cualesquiera. Entonces, si $\xi_{\lambda_1}(A, B)(z) < 0.5$, $z \in U$ se tiene $\xi_{\lambda_1}(A, B)(z) \ge \xi_{\lambda_2}(A, B)(z)$.

Demostración: Sea $\xi_{\lambda_1}(A, B)(z) < 0.5$, $z \in U$, por tanto $\inf_{x \in U} \{\min(1, \lambda_1(B(x)) + \lambda_1(1 - A(x+z)))\} < 0.5$. Esto implica que el subconjunto $D \subseteq U$ tal que

$$D = \{x / \lambda_1(B(h)) + \lambda_1(1 - A(h+z)) < 0.5\} \text{ es } \neq \varnothing$$

En consecuencia, (por la hipótesis $\xi_{\lambda_1}(A, B)(z) < 0.5$),

$$\xi_{\lambda_1}(A, B)(z) = \inf_{x \in U} \{\min(1, \lambda_1(B(x)) + \lambda_1(1 - A(x+z)))\} =$$

$$= \inf\{ \min(1, \lambda_1(B(x)) + \lambda_1(1 - A(x+z))) / \; x \in D \}$$

Además, de $\lambda_1(B(x)) + \lambda_1(1 - A(x+z)) < 0.5$ para todo $x \in D$ deducimos que $\lambda_1(B(x)) < 0.5$ y $\lambda_1(1 - A(x+z)) < 0.5$ y, (puesto que para cualquier λ se verifica la equivalencia $p < 0{,}5 \Leftrightarrow \lambda(p) \ge 0.5$), en consecuencia

$$B(x) \ge 0.5 \text{ y } 1 - A(x+z) \ge 0.5 \text{ para todo } x \in D.$$

Se concluye de estas últimas desigualdades que

$\lambda_1(B(x)) \ge \lambda_2(B(x))$ y $\lambda_1(1 - A(x+z)) \ge \lambda_2(1 - A(x+z))$ $\forall x \in D$ y de las desigualdades del mismo sentido obtenidas a partir de la suma de las anteriores, junto a las hipótesis exigidas deducimos finalmente que

$$\xi_{\lambda_1}(A, B)(z) = \inf\{\min(1, \lambda_1(B(x)) + \lambda_1(1 - A(x+z))) / \; x \in D\} \ge$$

$$\inf\{\min(1, \lambda_2(B(x)) + \lambda_2(1 - A(x+z))) / \; x \in D\} \ge$$

$$\inf_{x \in \mathbf{U}} \{\min(1, \lambda_2(B(x)) + \lambda_2(1 - A(x+z)))\} = \xi_{\lambda_2}(A, B)(z). \quad .$$

Teorema 2: Dados ξ_{λ_1} y $\xi_{\lambda_2} \in E$ tal que $\xi_{\lambda_1} \trianglelefteq \xi_{\lambda_2}$, A, B$\in F(U)$ verificando $\forall y \in \mathbf{U}$:

$$\lambda_2(1-A(x)) \geq \lambda_1(1-A(y)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\}$$

Entonces se tiene $\xi_{\lambda_1}(A, B)(x) \leq \xi_{\lambda_2}(A, B)(x) \quad \forall x \in \mathbf{U}$.

Demostración: Vamos a probar primero que

$$\lambda_1(B(p)) \leq \lambda_2(B(p)) + \sup_{x \in \mathbf{U}} \{\lambda_1(B(x)) - \lambda_2(B(x))\} \quad \forall p \in \mathbf{U}:$$

- Si $B(p) \leq 0.5 \ \forall p \in \mathbf{U}$, entonces $\sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\} \leq 0$, y resulta que

$$\lambda_1(B(p)) \leq \lambda_2(B(p)) + \sup_{x \in \mathbf{U}} \{\lambda_1(B(x)) - \lambda_2(B(x))\} \quad \forall p \in \mathbf{U}$$

- Si $\exists\, k \in \mathbf{U}$ tal que $B(k) > 0.5 \Rightarrow$ $\lambda_1(B(k)) \geq \lambda_2(B(k)) \Rightarrow \lambda_1(B(k)) - \lambda_2(B(k)) \leq$
- $\sup_{x \in \mathbf{U}} \{\lambda_1(B(x)) - \lambda_2(B(x))\} \Rightarrow \lambda_1(B(k)) \leq$ $\lambda_2(B(k)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\}$.

Al ser

$$\lambda_2(1-A(y)) \geq \lambda_1(1-A(y)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\}$$

$\forall y \in \mathbf{U}$, para cualquier x perteneciente a **U** podemos escribir

$$\lambda_2(1-A(z+x)) \geq$$

$$\lambda_1(1-A(z+x)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\} \quad \forall z \in \mathbf{U},$$

y por tanto $\lambda_2(1-A(z+x)) + \lambda_2(B(z)) \geq$

$$\lambda_1(1-A(z+x)) + \lambda_2(B(z)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\} \quad \forall z \in \mathbf{U},$$

y puesto que

$$\lambda_2(B(z)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\} \geq \lambda_1(B(z)),$$ resulta que

$$\lambda_2(1-A(z+x)) + \lambda_2(B(z)) \geq \lambda_1(1-A(z+x)) + \lambda_1(B(z)) \ \forall z \in \mathbf{U},$$

y esto implica que $\xi_{\lambda_2}(A, B)(x) \geq \xi_{\lambda_1}(A, B)(x) \quad \forall x \in \mathbf{U}$ tal que

$$\lambda_2(1-A(y)) \geq \lambda_1(1-A(y)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\}$$

$\forall y \in \mathbf{U}$.

Corolario 3: Dadas dos erosiones borrosas ξ_{λ_1} y ξ_{λ_2} tal que $\xi_{\lambda_1} \trianglelefteq \xi_{\lambda_2}$, entonces $0.5 \leq \xi_{\lambda_1}(A, B)(x) \leq \xi_{\lambda_2}(A, B)(x)$, para todo $x \in \mathbf{U}$ tal que

$$\lambda_2(1-A(y)) \geq \lambda_1(1-A(y)) + \sup_{h \in \mathbf{U}} \{\lambda_1(B(h)) - \lambda_2(B(h))\}$$ y

$$A(y) > 0.5 \quad \forall y \in \mathbf{U}.$$

Demostración: Es una consecuencia del teorema 3.

Estamos interesados en señalar la relación que existe entre el conjunto $(F_g \prec)$ de las funciones g mencionadas en la introducción con el orden inducido puntualmente, el conjunto ordenado $(F_\lambda \prec)$ de las funciones λ obtenido a partir del anterior y finalmente, el conjunto ordenado $(E \trianglelefteq)$ de las erosiones definidas con estas λ.

Sea F_g el conjunto de las funciones

$$g: [0, 0.5] \longrightarrow [0.5, 1]$$

tales que: $g(0) = 1$, g estrictamente decreciente y g continua en el punto 0.

Teorema 4: F_g es un retículo con el orden inducido $\prec$ a partir del orden en $[0, 0.5]$:

$g_1 \prec g_2$ si y sólo si $g_1(x) \leq g_2(x) \ \forall x \in [0, 0.5]$.

- Demostración:

Sean g_1, g_2 y g_3 funciones pertenecientes a Fg, entonces:

- $(g_1 \wedge g_2)(0) = 1$ y $g_1 \vee g_2\,(0) = 1$
- $(g_1 \wedge g_2)(x)$ y $(g_1 \vee g_2)(x)$ son continuas en el punto $x=0$ ya que g_1, g_2 ,son continuas en 0 y las funciones max y min are continuos.
- $(g_1 \wedge g_2)(x)$ y $(g_1 \vee g_2)(x)$ son estrictamente decrecientes puesto que g_1 y g_2 son estrictamente decrecientes.

Por tanto $g_1 \wedge g_2$ y $g_1 \vee g_2$ son funciones pertenecientes a Fg y (Fg, $\prec$) es un retículo.

Nótese que este retículo no es completo, pues si se considera la familia de funciones $\{g_n \mid g_n(p) = 1 - \frac{1}{n}p + 1, \ \forall p \in [0, 0.5], n \in N\}$ es claro que su supremo es $g(p) = 1$ que no es estrictamente decreciente.

Sea F_λ el conjunto de las funciones λ de [0, 1] en [0, 1] tal que $\lambda(p) = \begin{cases} g(p) & \text{si } p \in [0, 0.5] \\ 1 - g(1-p) & \text{si } p \in (0.5, 1] \end{cases}$, $g \in F_g$,, Se verifica:

Corolario 5: F_λ es un retículo con la relación de orden $\lambda_1 \prec \lambda_2$ si y sólo si $\lambda_1(p) \leq \lambda_2(p) \ \forall p \in [0, 0.5]$.

Demostración: Prueba análoga a la del Teorema 4.

Análogamente se puede ver que el retículo $(F_\lambda, \prec)$ no es completo.

Teorema 6.- Los retículos $(F_\lambda, \prec)$ y $(F_g, \prec)$ son retículos isomorfos.

Demostración:

- Hemos visto que F_g y F_λ son conjuntos ordenados y Es inmediato que existe a aplicación biyectiva entre F_g y F_λ por construcción de la función λ.

- Es isomorfismo de retículos ya que para $g_1 \leq g_2$ por construcción de las respectivas λ_1 y λ_2 tenemos que $\lambda_1 \leq \lambda_2$ y recíprocamente. .

Teorema 7: El conjunto de las erosiones borrosas E con la relación de orden $\trianglelefteq$ es un retículo.

Demostración: Dadas las erosiones ξ_{λ_1} y ξ_{λ_2} asociadas respectivamente a g_1 y a g_2, las erosiones asociadas a $g_1 \vee g_2$ y a $g_1 \wedge g_2$ son respectivamente el supremo y el ínfimo de las iniciales.

REFERENCIAS

[1] S. Divyendu y E. R. Dougherty, Characterization of Fuzzy Minkowski Algebra, SPIE Vol. 1769 Image Algebra and Morphological Image Processing III (1992) 59-69.

[2] S. Divyendu y E. R. Dougherty, Fuzzification of Set inclusion, SPIE Vol. 1708 Applications of Artificial Intelligence X: Machine Vision and Robotics (1992) 440-449.

[3] S. Divyendu y E. R. Dougherty, Fuzzification of set inclusion: Theory y applications, Fuzzy Sets and Systems 55 (1993) 15-42.

[4] S. Divyendu y E. R. Dougherty, Fuzzy Mathematical Morphology, Journal of Visual Communication and Image Representation, Vol. 3, No. 3, September 1992, 286-302.

[5] D. Dubois y H. Prade, Fuzzy Sets y Systems: Theory and Applications, Mathematics in Science y Engineering, (Academic Press, New York 1980).

[6] N. Frago, Morfología Matemática Borrosa basada en operadores generalizados de Lukasiewicz: Procesamiento de imagenes. Ph.D Thesis, Universidad Pública de Navarra, Spain 1996.

[7] P, Burillo,. N. Frago, R. Fuentes, Generación de Morfologías Matematicas Borrosas, Actas del VIII Congreso Español sobre Tecnologías y Lógica Fuzzy, Pamplona 1998.

[8] J. Serra, Image Analysis and Mathematical Morphology, Vol. 1 y 2, Academic Press, London 1982.

DECISION SUPPORT SYSTEM BASED ON FUZZY KNOWLEDGE APPLIED TO A SOFTWARE FACTORY

F. Chamorro
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid
felixcha@inf.uc3m.es

F. García
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid
fgarcia,@inf.uc3m.es

A. Sanchis
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid
masm@inf.uc3m.es

J. M. Molina
Dpto de Informática
EPS, U. Carlos III
Avda. Universidad 30
28911 Leganés Madrid
molina@ia.uc3m.es

Abstract

Development environments to DSS are evolving to open and distribute architectures. The main features of these new architectures are: Data Base Query Tools in clients, fast developments and better graphics interfaces. DSS developers need a set of functionalities as multidimensional data view, variable capacity DW, set of complex criteria in order to filter the required information access, quick queries, minimum reorganisation when the data model is modified, etc. In this work a fuzzy knowledge based system to make decisions inside a DSS system that supports the planning and production control in a software factory. The fuzzy system is developed to evaluate the appropriateness of different platforms to construct a software product. From the platform specifications given in the product definition, the fuzzy system calculates the appropriateness of each existing platform to develop the product.
Keywords: Fuzzy Decision, Knowledge Based Systems, Groupware, Decision Support Systems

1 INTRODUCTION

Two main business features are the base for the analysis and construction of a DSS tool: the strategic planning of the organization and the structure of the information system in a modern and competitive business. Software tools make easier the access to the necessary information The data obtained ought to be conveniently structured in time and following pre-defined patterns.

Knowledge and experience constitute the only truly durable resource of the whole organization. The ability to capture and manage such knowledge is necessary for the survival of any organization [4]. Recently, a new type of software has been developed, which under the title of groupware, allows to create, manage, share and take full advantage of resources in the same way that knowledge and experience light in a collective work [3]. These applications are characterized by:

- Considering communications as the central axis of decision and business management
- Contributing in a decisive way to the automation of processes that take place in the organization [2].

The introduction of strategic systems in business substantially improves the quality of the information used in processes. It drastically reduces the time spent in executing these processes and makes access to the information easier. The scope of these systems is:

- applications for sales management and accounting
- applications for product development
- quality control applications or applications for customer service

The information could be far from the point where these decisions are taken and, besides, business organisations keep their information distributed through different databases. These characteristic are relevant for DSS applications. Networks of information and communication allow the integration of DSS inside distributed strategic business systems [8]. The communication allows to capture the information for taking the decision and to distribute the results. The current network systems LAN/WAN and networks based on Internet are governed by groupware with workflow applications which provides the information, as well as carrying out the distribution of the already requested relevant information [3].

Another aspect that should not be overlooked is the usefulness that knowledge based systems can bring to these applications [1]. In general, at the level of support for decisions made in business, tools exist which could influence in a very positive way the decision making process. These tools belong to the field of knowledge engineering such as the aforementioned [6]. This article influences and aims at contributing to the development of this field of information technology in relation to the levels of tactical planning and strategy of the organizations.

The basic objective aimed at in the design of our prototype proposes to join the advantages and functions that the strategic systems offer with those that knowledge engineering can provide for Expert Systems (ESs) [7]. A prototype has been obtained that makes the choice of decisions to be made by the components of a department group or business dedicated to the manufacture of

software. This application decides the most convenient platform (including hardware, software, cost, etc..) for a software development.

2 DESIGN AND DEVELOPMENT OF DSS WITH EXPERT KNOWLEDGE

The common objectives of pursuing an information technology tool of this type can be a innumerable from these we will select some of the fundamentals in the design of the prototype divided into five levels:

- Capturing information: connecting existing databases in the business through the Data Warehouse, using the transmission system of data which are used in the business and distribution of information from remote machines.
- Functional and generation of reports: generating their own stores of intervening information data marts, employing easy to use screens, interpreting and modifying, improving the execution of the application, making them easier to use, configuration of the different types of reports and flexibility and speed in the management of information.
- Distribution of information: connection to groupware, working information with workflow, connection to other applications of communication and electronic mail and base of sending and receiving reports.
- Projects control: Through those components from Lotus Notes 4.1 the integration with applications for planning and programming projects should be permitted, like Microsoft Project or CA Superproject, in order to offer the user all the information and functions that both types of application offer.
- Generals: ease of maintenance, possibility of dynamic level evaluation and bringing things up to date customer architecture.

The design and development of the prototype DSS system is based on the general objective of the system. This system supports the planning and production control in a software factory. With a system like this many advantages can be obtained [6] and the control of projects can be improved, helping the director of production to solve internal and external problems caused by resources and customers which occur in the factory in the development of plan and projects for the manufacture of software.

This prototype will help solve any problems which might occur in the course of the planning and production of software in the software factory. The system that will be developed will be known as ISESD (Intelligent System for Events during Software Development) and Figure 1 shows the main window of application. The person in charge of planning in the factory must confront the situations from the point of view of performance and from the perspective of achieving production targets and revision in the development and maintenance of software.

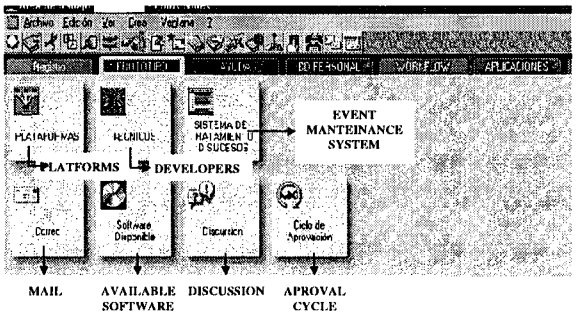


Figure 1: ISESD Interface.

3 PLATFORM FUZZY EVALUATION

To automate the pursued human-like platform evaluation process, the first step should be to select adequate descriptions of a platform, "analogous" to those formulated by humans when interpreting the requirements of software product. Artificial Intelligence representation schemes best suit these requirements. A generic platform will be represented as a set of characteristic, grouped in general features about platform as hardware, software, cost, availability, etc... This information could be represented by $(F_1,F_2,...,F_n)$, a set of values characterizing a platform; each $f_i \in F_i$, being F_i the range of possible numerical values of f_i.

Platform characteristics can then be related with a metric distance in the space of possible values. Given a numerical distance, d_i, between f_i (values of two different platforms, respectively), D_i is defined as the range of all possible values of d_i. To better cope with the intrinsic uncertainty of the perceived features, numerical values of the distances d_i will be mapped into qualitative symbolic descriptions. Thus, the computed distances will be transformed, through a fuzzification process, into linguistic variables. For each d_i, a linguistic variable, Ld_i, is introduced together with its set of values $\{ld_{i1},ld_{i2},...,ld_{ic_i}\}$, each one labelling a fuzzy subset in D_i (its meaning). The fuzzification operation applied to the numerical distances d_i, will result in their transformation into a fuzzy singleton [9], fuzzy subset whose membership function is the Kronecker delta, $\delta(d-d_i)$, in D_i.

A fuzzy relational algorithm (FRA) [9] will store the knowledge required to obtain the appropriateness of a platform with the desired platform, based on the linguistic similarities between the features of each one. The FRA will be composed by a finite set of fuzzy conditional statements of the form IF $AND^n_{i=1}(Ld_i$ is $ld_{ij_i})$ THEN

(APPR is ls_k), whose antecedents are conjunctions of fuzzy statements about the linguistic variables Ld_i (i=1,2,...,n), and their consequents fuzzy statements about APPR, linguistic appropriateness of the platform, whose set of values is $\{ls_1, ls_2, ..., ls_p\}$. The well known Mamdani implication [5] has been chosen to assign their meaning to these fuzzy conditional statements.

The compositional rule of inference (CRI) [8], (approximate extension of the rule of *modus ponens*) has been adopted as the inference mechanism to obtain the fuzzy subset induced in APPR (through each conditional statement of the FRA), by the fuzzy statement of the form $AND^n_{i=1}(Ld_i \text{ is } ld_{ir_i})$, linguistic description of the distance among platform features. The meaning of APPR will be the intersection of the intermediate meanings resulting from the application of the CRI to each conditional statement of the FRA (min of all the induced consequent membership functions). Finally, the adopted defuzzification process, applied to the final meaning of APPR, will be the traditional Centre of Gravity procedure [10].

4 FUZZY SYSTEM DEVELOPED

The final goal of this work is to develop a knowledge based system that using the fuzzy logic help project manager of a software factory to choose the more adequate platform for the accomplishment of a project. Each platform is defined by four numerical concepts that correspond, in order: hardware, software, availability and cost.

To verify the validity of this system, it is necessary to compare the results obtained with the fuzzy system with the results obtained using euclidean distance between the needed platform and the existing standard platforms.

4.1 INPUT VARIABLES

Four variables are considered and corresponding with the distances between the needed platform and the existing platforms: the distance between the hardware characteristics, represented by V1; the distance between the software characteristics, V2; the distance between the availability characteristics, V3; and the distance between the cost characteristics, V4. Membership functions of the input variables are the same type and each one is classified in three membership functions:

1.- Near (N): The distance is so small that the corresponding value of the standard platform is approximately equal to that of the needed one.

2.- Medium (M): The standard platform is approximated acceptably to the needed one.

3.- Far (F): The values correspond at greater discrepancy level, or separation, between platforms.

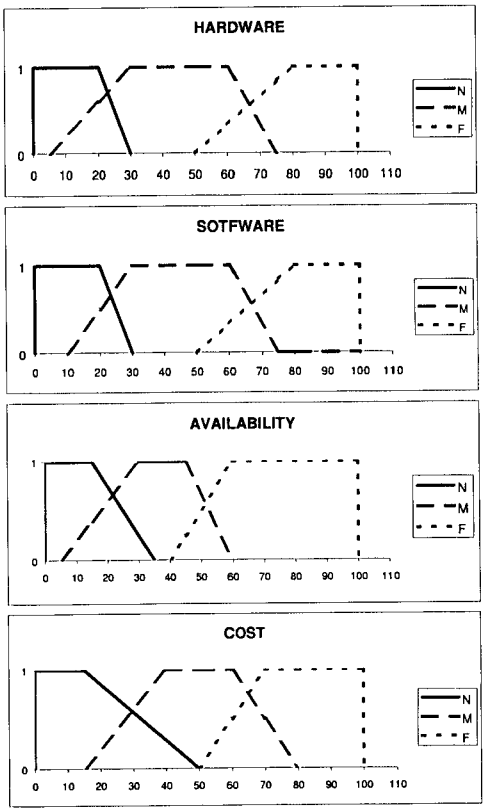


Figure 2: Membership Functions of Input Variables

4.2 OUTPUT VARIABLE.

An ouput variable (S) represents the degree of proximity of the different existing standard platforms to the needed platform for the project. The range of possible values for the output variable is 0 to 200, since this is also the range for the euclidean distance. This definition allows to compare the distances obtained by the two methods. To obtain a more detailed classification of the analyzed platforms, five membership functions are used. When distance is around zero the analyzed platform is near to the needed one, on the other hand, when is aorund 200 the analyzed platform is far from the needed one. The membership functions (Very Near, Near, Medium, Far, Very Far) are represented in Figure 3.

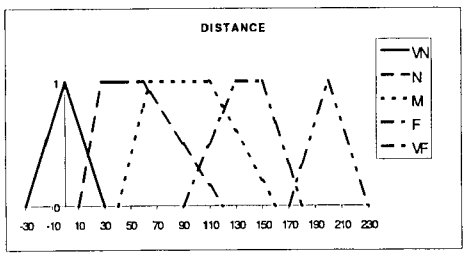


Figure 3: Membership Functions of output variable

4.3 RULES

In Figure 4, the resultant rules of the proposed system are shown.

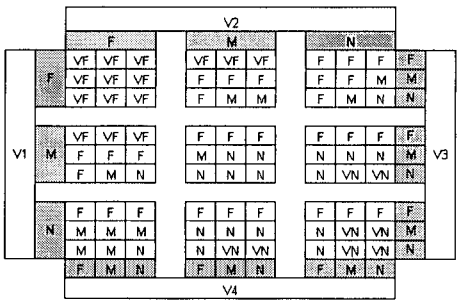


Figure 4: Rules of the system

Unfavorable situations have been taken into account, for example: a great distance in the availability of the platform in the requested time that makes impossible its use, a great distance in hardware or in software that gives a very high value. On the other hand, differences in the cost are considered less important, since this aspect could be almost always assumed.

5 RESULTS AND CONCLUSIONS

As it has been explained previously, the software factory has a data base with the values of the existing platforms. Some of these values are updated continually. As is the case of the availability that changed in function of the use of the platforms. The results presented in the Figure 5, 6 and 7 are obtained using the available values stored in the data base at the moment of accomplishing the experiments In Figure 5, the distance between the different existing platforms in the factory has been calculated. The white cells are the euclidean distance and the shadow ones are the fuzzy distances.

	UNIX	LINUX	NETWARE	WIN NT	WIN 9X	WIN 3X	MAC	AS400
UNIX	******	21,65	15,96	0	48,71	65,96	0	35,63
LINUX	40,69	******	44,02	39,87	30,01	63,64	33,36	107,37
NETWARE	22	54,44	******	0	41,42	65,04	0	43,05
WIN NT	16,79	43,89	13,93	******	40,06	56,42	0	44,88
WIN 9X	59,57	51,96	59,26	48,15	******	7,05	51,48	119,43
WIN 3X	88,19	72,95	85,88	76,27	30,45	******	56,42	157,55
MAC	31,83	40,36	27,77	23,9	52,79	76,02	******	37,44
AS400	50,06	83,37	46,2	56,44	104,36	131,08	59,82	******

Figure 5: Distance between existing platforms

In Figure 6, the distance between the existing platforms and the platform needed for a small bussiness is showed.

UNIX	LINUX	NETWARE	WIN NT	WIN 9X	WIN 3X	MAC	AS400
42,31	25,96	32,16	35,18	0	30,58	42,31	96,96
60,07	45,43	60,3	49,21	14,28	30,94	47,34	103,61

Figure 6: Distance for a small bussiness

In Figure 7, the distance between the existing platforms and the platform needed for a high bussiness is showed.

UNIX	LINUX	NETWARE	WIN NT	WIN 9X	WIN 3X	MAC	AS400
0	25,56	34,53	27,44	56,42	78,88	26,53	35,04
19,87	29,65	35,85	28,3	61,46	88,42	28,25	57,26

Figure 7: Distance for a high bussiness

From results showed in Figures 5, 6 and 7 following conclusions could be summarised:

1) For values under 100, the distance obtained with the fuzzy system are always less than the one obtained by euclidean distance. Therefore, the fuzzy system evaluates better the proximity between the characteristics of the analysed platforms and the requiered ones.
2) When the value of the distance is over 100, the fuzzy distance is greater to the euclidean one what means that from a value in middle of range, the analysed platforms could not be used in the project.
3) The improvement of the fuzzy system relies on the different evaluation that could make for each concept. In the euclidean distance all the components have the same importance.

References

[1] Adelman, L. (1992), *Evaluating Decision Support and Expert Systems*. New York, NY: John Wiley & Sons.
[2] Bankos, J. Y., y Treacy, M. E. (1986), *Information Technology and Corporate Strategy: A Research Perspective*. MIS Quartely..
[3] Coleman, D. and Khanna, R. (1995), *Groupware: Technology and Applications*. Englewood Cliffs, NJ: Prentice-Hall PTR..
[4] Kimball, R. (1996), *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. New York: John Wiley & Sons.
[5] Mamdani H., "Aplication of Fuzzy Logic to Aproximate Reasoning Using Linguistic Synthesis", IEEE Trans. on Computers, Nov/Dic. 1988.
[6] Mockler, R. J. (1989), *Knowledge-based Systems for Strategic Planning*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.
[7] Olson, D. L. and Courtney, J. F Jr. (1992), *Decision Support Models and Expert Systems*. New York: Macmillan.
[8] Young, L. F. (1989), *Decision Support and Idea Processing Systems*. Dubuque, IA: Wm. C. Brown.
[9] Zadeh L. A., "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes", Trans. on SMC, Jan. 1973.
[10] Zimmermann H.J., *Fuzzy Sets, Uncertainty and Information*, Kluwer 1985.

A SELECTION METHOD BASED ON THE 2-TUPLE LINGUISTIC REPRESENTATION MODEL FOR DECISION-MAKING PROBLEMS WITH MULTI-GRANULARITY LINGUISTIC INFORMATION

F. Herrera[a], **Luis Martínez**[b]
[a]Dept. of Computer Science and A.I. University of Granada, 18071 - Granada, Spain. e-mail: herrera@decsai.ugr.es
[b] Dept. of Computer Science. University of Jaén, 23071 - Jaén, Spain. e-mail: martin@ujaen.es

Abstract

The use of linguistic information in decision-making problems implies in most cases the need for using fusion processes to obtain aggregated values that summarize the input information. An important limitation of the fuzzy linguistic approach appears when fusion processes are applied to problems in which the linguistic information is assessed in term sets with different granularity of uncertainty, this type of information is denoted as multi-granularity linguistic information. In this contribution, taking as the base the 2-tuple fuzzy linguistic representation model and its computational technique, we shall present a method for easily dealing with multi-granularity linguistic information in fusion processes.

Keywords: Linguistic variables, fusion processes, granularity of uncertainty, multi-granularity linguistic information, decision-making.

1 INTRODUCTION

In some occasions,we can find decision-making problems that present several sources of information to qualify a phenomenon. When these phenomena present quantitative aspects they can be assessed by means of precise numerical values, however when the aspects presented by the phenomena are qualitative, then it may be difficult to qualify them using precise values. So, the use of the fuzzy linguistic approach [8] has shown itself as a good choice to model these phenomena, due to the fact that it represents qualitative aspects with qualitative terms by means of linguistic variables. The use of the fuzzy linguistic approach implies computing with words (CWW). In the specialized literature, three different linguistic computational techniques that provide linguistic operators for CWW can be found [1, 2, 5].

An important aspect when the fuzzy linguistic approach is used, is to determine the "granularity of uncertainty", i.e., the cardinality of the linguistic term set. Depending on the uncertainty degree held by a source of information qualifying a phenomenon, the term set will have more or less terms. Then, in those problems with several sources of information each one could express its knowledge by means of linguistic term sets with a different granularity of uncertainty from the other ones. We shall denote this type of information as multi-granularity linguistic information.

In decision-making problems with multi-granularity linguistic information the fuzzy linguistic approach together with the linguistic computational techniques introduced in [1, 2] present an important limitation because in these computational methods, neither a standard normalization process nor fusion operators are defined for this type of information. Therefore, it is highly complex to solve this type of problems using these methods and the results obtained are expressed in domains far removed from the initial ones [3, 4].

The aim of this contribution is to present an easy selection model for decision-making problems with multi-granularity linguistic information using as a base the 2-tuple fuzzy linguistic representation model [5, 6], together with the multi-granularity linguistic information fusion ideas presented in [3, 4]. So, we shall present a selection model that obtains the solution set of alternatives according to the following two steps:

1. *A fusion process for multi-granularity linguistic information based on the 2-tuple representation,* for obtaining collective values for each alternative.

2. *A selection process,* for obtaining the solution set.

2 PRELIMINARIES

Here, we present the scheme of a multi-expert decision-making (MEDM) problem with multi-granularity linguistic information and different methods for CWW.

2.1 MEDM PROBLEM

An MEDM problem can be defined as follows. Let $X = \{x_1, x_2, \ldots, x_n\}$ $(n \geq 2)$ be a finite set of alterna-

tives to be qualified according to a finite set of experts $P = \{p_1, p_2, \ldots, p_m\}$ $(m \geq 2)$. Each expert p_i provides a linguistic performance value μ^{ij} for each alternative x_j. Given that we shall deal with multi-granularity linguistic term sets in decision-making problems, we assume that each expert p_i may use a different linguistic term set S_i to express the performance values. The linguistic term sets $\{S_i, \forall i\}$ may have a different granularity and/or semantics. Therefore, for each expert p_i, the performance profile of the alternatives is defined as a linguistic fuzzy choice subset defined over X and assessed linguistically on S_i:

$$p_i \longrightarrow (\mu^{i1}, \ldots, \mu^{in})$$
$$\mu^{ij} \in S_i \quad S_i = \{s_0^i, \ldots, s_{g_i}^i\} \quad i \in \{1, \ldots, m\}$$

where $g_i + 1$ is the cardinality of S_i.

2.2 LINGUISTIC COMPUTATIONAL METHODS

The linguistic variables are used in processes of CWW that imply their fusion, aggregation, comparison, etc. To perform these computations there are three models in the literature. (i) The model based on the Extension Principle, (ii) the symbolic one and (iii) the model based on the 2-tuple fuzzy linguistic representation. Here we briefly review the first two models and we shall describe in depth the last one.

1. *The linguistic computational methods based on the Extension Principle [2].* These methods use the extended arithmetic, based on the Extension Principle, on the membership functions associated to the linguistic terms to make linguistic computations.

2. *The linguistic computational symbolic models [1].* These methods do not use the membership functions of the labels to perform the computations, but they use the order index and properties of such linguistic assessments to make direct computations on labels.

3. *The 2-tuple Fuzzy Linguistic Representation Model [5, 6].* It is based on symbolic methods and takes as the base of its representation the concept of Symbolic Translation.

Definition 1. *The Symbolic Translation of a linguistic term $s_i \in S = \{s_0, ..., s_g\}$ is a numerical value assessed in $[-.5, .5)$ that supports the "difference of information" between a counting of information β assessed in $[0, g]$ obtained after a symbolic aggregation operation (acting on the order index of the labels) and the closest value in $\{0, ..., g\}$ that indicates the index of the closest linguistic term in S (s_i).*

From this concept we develop a linguistic representation model which represents the linguistic information by means of 2-tuples, (r_i, α_i), $r_i \in S$ and $\alpha_i \in [-.5, .5)$. r_i represents the linguistic label center of the information and α_i is the Symbolic Translation.

This linguistic representation model defines a set of functions to make transformations among linguistic terms, 2-tuples and numerical values.

Definition 2. *Let $s_i \in S$ be a linguistic term, then its equivalent 2-tuple representation is obtained by means of the function θ as:*

$$\theta : S \longrightarrow (S \; x \; [-0.5, 0.5))$$
$$\theta(s_i) = (s_i, 0)/s_i \in S$$

Definition 3. *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and $\beta \in [0, g]$ a value supporting the result of a symbolic operation, then the 2-tuple that expresses the equivalent information to β is obtained with the following function:*

$$\Delta : [0, g] \longrightarrow Sx[-0.5, 0.5)$$

$$\Delta(\beta) = \begin{cases} s_i & i = round(\beta) \\ \alpha = \beta - i & \alpha \in [-.5, .5) \end{cases}$$

where round is the usual round *operation, s_i has the closest index label to "β" and "α" is the value of the symbolic translation.*

Definition 4. *Let $S = \{s_0, ..., s_g\}$ be a linguistic term set and (s_i, α) be a linguistic 2-tuple. There is always a function Δ^{-1}, such that, from a 2-tuple it returns its equivalent numerical value $\beta \in [0, g]$.*

$$\Delta^{-1} : Sx[-.5, .5) \longrightarrow [0, g]$$
$$\Delta^{-1}(s_i, \alpha) = i + \alpha = \beta$$

Together with the fuzzy linguistic 2-tuple representation model a wide range of 2-tuple aggregation operators were developed in [5], which are neccesary for the development of our fusion method in order to combine the information.

3 A FUSION PROCESS FOR MULTI-GRANULARITY LINGUISTIC INFORMATION BASED ON THE 2-TUPLE REPRESENTATION MODEL FOR MEDM PROBLEMS

We want to obtain for each alternative, x_j, a collective value expressed by means of a linguistic 2-tuple. To do so, we present the following fusion process:

1. **Making the information uniform** (Normalization process). In this step the multi-granularity linguistic input information is unified into "fuzzy sets" in a Basic Linguistic Term Set (BLTS).

2. **Transforming fuzzy sets into 2-tuples.** Here we shall transform the fuzzy sets in the BLTS into 2-tuples based on the symbolic translation assessed in the BLTS.

3. **Fusion of 2-tuples.** Once the performance values, μ^{ij}, are expressed by 2-tuples assessed in the BLTS, we shall apply a 2-tuple fusion operator to obtain collective performance values.

4. **Backward step.** The 2-tuples obtained by the fusion method are assessed in the BLTS, it can be distant from the expression domains used by the sources of information. Therefore, it may be interesting to offer the option to make an approach of the collective values to the initial domains for a better comprehensiveness. This step is not neccesary, simply convenient.

Subsequently, we shall present in depth each step.

3.1 MAKING THE INFORMATION UNIFORM

With a view to manage the information we must make it uniform, i.e., the multi-granularity linguistic information must be transformed into a unified linguistic term set, called BLTS and denoted as S_T. Before defining a transformation function into this BLTS, S_T, we have to decide how to choose S_T. We consider that S_T must be a linguistic term set which allows us to maintain the uncertainty degree associated to each expert and the ability of discrimination to express the performance values. With this goal in mind, we look for a BLTS with the maximum granularity. We take into consideration two possibilities:

1. When there is only one term set with the maximum granularity, then, it is chosen as S_T.

2. If we have two or more linguistic term sets with maximum granularity then, S_T is chosen depending on the semantics of these linguistic term sets, finding two possible situations to establish S_T:

(a) All the linguistic term sets have the same semantics, then S_T is any one of them.

(b) There are some linguistic term sets with different semantics. Then, S_T is a term set with a larger number of terms than the number of terms that a person is able to discriminate (normally 11 or 13, see [7]). We define a BLTS with 15 terms simmetrically distributed.

Once S_T has been selected, we define a transformation function which transforms each linguistic value into a fuzzy set in S_T.

Definition 5 [3]. *Let $A = \{l_0, \ldots, l_p\}$ and $S_T = \{c_0, \ldots, c_g\}$ be two linguistic term sets, such that, $g \geq p$. Then, a multi-granularity transformation function, τ_{AS_T} is defined as,*

$$\tau_{AS_T} : A \longrightarrow F(S_T)$$
$$\tau_{AS_T}(l_i) = \{(c_k, \alpha_k^i) \ / k \in \{0, \ldots, g\}\}, \ \forall l_i \in A$$
$$\alpha_k^i = \max_y min\{\mu_{l_i}(y), \mu_{c_k}(y)\}$$

being $F(S_T)$ the set of fuzzy sets defined in S_T, $\mu_{l_i}(y)$ and $\mu_{c_k}(y)$ are the membership functions of the fuzzy sets associated to the terms l_i and c_k, respectively.

We shall denote each $\tau_{S_i S_T}(\mu^{ij})$ as r^{ij}, and represents each fuzzy set of performance, r^{ij}, by means of its respective membership degrees, i.e.,

$$r^{ij} = (\alpha_0^{ij}, \ldots, \alpha_g^{ij}).$$

3.2 TRANSFORMING FUZZY SETS INTO 2-TUPLES

So far, we have unified the multi-granular linguistic information into fuzzy sets in S_T. The fuzzy sets are complex to manage. Therefore, we shall use the 2-tuple fuzzy linguistic representation model to represent this information. To do so, we shall define the function χ that computes a value $\beta \in [0, g]$ that supports the information in the fuzzy set $\tau_{S_i S_T}(\mu^{ij})$.

Definition 6. *Let $\tau_{S_i S_T}(l_i) = \{(c_0, \alpha_0^i), \ldots, (c_g, \alpha_g^i)\}$ be a fuzzy set that represents a linguistic term $l_i \in S_i$ in S_T. We shall obtain a numerical value, that supports the information of the fuzzy set, assessed in the interval $[0, g]$ by means of the following function:*

$$\chi : F(S_T) \longrightarrow [0, g]$$
$$\chi(\tau_{S_i S_T}(l_i)) = \frac{\sum_{j=0}^{g} j\alpha_j^i}{\sum_{j=0}^{g} \alpha_j^i} = \beta$$

This value β is easy to transform into a linguistic 2-tuple using the function Δ (Definition 3). Therefore, we have unified the input information with linguistic 2-tuples assessed in S_T transforming the fuzzy sets, r^{ij}, by means of the functions χ and Δ:

$$\Delta(\chi(\tau_{S_i S_T}(\mu^{ij}))) = \Delta(\chi(r^{ij})) = (s_k, \alpha)^{ij}$$

with $s_k \in S_T$ *and* $\alpha \in [-.5, .5)$.

3.3 FUSION OF 2-TUPLES

Now the performance values, μ^{ij}, are modeled by means of linguistic 2-tuples assessed in S_T, $(s_k, \alpha)^{ij}$, and our objective is to aggregate this information to obtain collective values for each alternative x_j. In [5] a wide range of 2-tuple linguistic aggregation operators were presented, therefore, to aggregate the 2-tuples, $(s_k, \alpha)^{ij}$, we shall choose one of these operators and apply it for combining the 2-tuples, obtaining as result an aggregated linguistic 2-tuple assessed in S_T. Formally, it can be expressed as:

$$FO((s_k, \alpha)^{1j}, ..., (s_k, \alpha)^{nj})) = (s_k, \alpha)^j$$

where FO is any 2-tuple fusion operator, and $(s_k, \alpha)^j$ is the collective value for the alternative, x_j.

3.4 THE BACKWARD STEP

This is an "*optional step*" in the fusion process. Depending on the problem we are dealing with, it may be that the aggregated 2-tuple assessed in S_T is expressed in a expression domain distant from the domains, S_i, used by the information sources. In these situations, to offer the possibility of making an approach to the initial expression domains, for improving the comprehensiveness of the results, might be appropiate. To accomplish the backward step we shall present:

1. *A new representation for the linguistic information.* Using 2-tuples based on the "degree of membership", i.e., 2-tuples whose first component is a linguistic label and the second one indicates the degree of membership of the counting of information represented in the linguistic term. The following function δ transforms a 2-tuple based on the symbolic translation into two 2-tuples based on the degree of membership supporting the same counting of information.

Definition 7. *Let (s_k, α) be a linguistic 2-tuple based on the symbolic translation with $s_k \in S_T$ and $\alpha \in [-.5, .5)$ whose equivalent numerical value is $\Delta^{-1}((s_k, \alpha)) = \beta$ (Definition 4) with $\beta \in [0, g]$. The function δ computes two 2-tuples based on the degree of membership, from the initial 2-tuple, that support the same counting of information:*

$$\delta : [0, g] \longrightarrow \{S_T x[0,1]\} x \{S_T x[0,1]\}$$
$$\delta(\beta) = \{(s_h, 1-\gamma), (s_{h+1}, \gamma)\}$$

where

$$h = trunc(\beta) \quad and \quad \gamma = \beta - h$$

trunc is the usual trunc operation.

2. *A process that obtains a 2-tuple (s_k^i, α), with $s_k^i \in S_i$ and $\alpha \in [-.5, .5)$ based on the symbolic translation, from two 2-tuples based on the degree of membership assessed in a different domain from S_i.*

(a) A matching procedure using $\tau_{S_T S_i}$ is applied to s_h *and* s_{h+1} obtaining two fuzzy sets on S_i.

$\tau_{S_T S_i}(s_h) = \{(s_0, \alpha_0^h), ..., (s_{g_i}, \alpha_{g_i}^h)\}$ *with* $r_h = (\alpha_0^h, ..., \alpha_{g_i}^h)$
$\tau_{S_T S_i}(s_{h+1}) = \{(s_0, \alpha_0^{h+1}), ..., (s_{g_i}, \alpha_{g_i}^{h+1})\}$
with $r_{h+1} = (\alpha_o^{h+1}, ..., \alpha_{g_i}^{h+1})$

(b) The fuzzy sets are converted into numerical values assessed in $[0, g_i]$,

$$\chi(\tau_{S_T S_i}(s_h)) = \chi(r_h) = \beta_h$$
$$\chi(\tau_{S_T S_i}(s_{h+1})) = \chi(r_{h+1}) = \beta_{h+1}$$

(c) To achieve our objective we need to obtain a value β assessed in $[0, g_i]$ that supports the same information as $(s_h, 1-\gamma)$, (s_{h+1}, γ). We have β_h *and* $\beta_{h+1} \in [0, g_i]$, that represent the information supportted by s_h *and* s_{h+1}, now we make a linear combination using the degrees of membership of these labels in their respective 2-tuples to obtain the value that we are looking for:

$$(\beta_h * (1-\gamma)) + (\beta_{h+1} * \gamma) = \beta \in [0, g_i]$$

where β represents the same information as the two 2-tuples based on the degree of membership. Then $\Delta(\beta)$ obtains the linguistic 2-tuple based on the symbolic translation assessed in S_i that we were looking for:

$$\Delta(\beta) = (s_k^i, \alpha)$$

This process will be carried out $\forall$ S_i.

4 SELECTION PROCESS

The objective of the decision process is to find a set of alternatives with the best ones. To do so, a selection process is applied to the collective preferences:

1. First, we select a choice degree.

2. The choice degree is applied to the collective values to rank them.

3. Finally, we select for the solution set the best alternatives according to the choice degree.

5 CONCLUDING REMARKS

In this paper we have presented a fusion method based on the 2-tuple fuzzy linguistic representation that allows us to easily deal with multi-granularity linguistic information in fusion processes.

REFERENCES

[1] M. Delgado, J.L. Verdegay and M.A. Vila, On Aggregation Operations of Linguistic Labels, *Int. J. of Intelligent Systems* **8** (1993) 351-370.

[2] R. Degani, G. Bortolan, The problem of Linguistic Approximation in Clinical Decision Making, *Int. J. of Approximate Reasoning* **2** (1988) 143-162

[3] F. Herrera, E. Herrera-Viedma, L. Martínez , A Fusion Approach for Managing Multi-Granularity Linguistic Terms Sets in Decision Making, *Fuzzy Sets and Systems*, (1999). To appear.

[4] F. Herrera, E. Herrera-Viedma, L. Martínez, Un problema de Selección para Problemas de Decisión con Múltiples Expertos e Información Linguística Multi-Granular, *Actas del VIII Congreso ESTYLF*, Pamplona, (1998) 333-338.

[5] F. Herrera, L. Martínez, A 2-tuple Fuzzy Linguistic Representation Model for Computing with Words, Dept. Computer Sciences and A.I., Granada University, TR#DECSAI-990102, 1999. (http://decsai.ugr.es/~herrera/ public.html)

[6] F. Herrera, L. Martínez, A Fuzzy Linguistic REpresentation Model based on a Symbolic Trnaslation, *Proceedings of EUROFUSE-SIC 99*, Budapest, (1999), 178-183.

[7] G.A. Miller, The Magical Number Seven or Minus Two: Some limits On Our Capacity of Processing Information, *Psychological Rev.* **63** (1956) 81-97.

[8] L.A. Zadeh, The Concept of a Linguistic Variable and Its Applications to Approximate Reasoning. Part I, *Information Sciences* **8** (1975) 199-249, Part II, *Information Sciences* **8** (1975) 301-357, Part III, *Information Sciences* **9** (1975) 43-80.

PROBABILITY OF INTUITIONISTIC FUZZY EVENTS AND THEIR APPLICATIONS IN DECISION MAKING

Eulalia Szmidt and Janusz Kacprzyk
Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, 01–447 Warsaw, Poland
E-mail: {szmidt, kacprzyk }@ibspan.waw.pl

Summary

A concept of intuitionistic fuzzy events and their probabilities is introduced. The solution is given by an interval and is consistent with the probability measure proposed by Zadeh for fuzzy sets.

The presented considerations seem to be crucial in decision making where imperfection of information is a rule. There are many aspects of information imperfection and among them uncertainty (randomness) and imprecision (fuzziness) are the most important. In this article we assume that imprecission is modelled by intuitionistic fuzzy sets and uncerainty is modelled by probability theory. Both types of information imperfection are discussed and expressed by common formulas.

Keywords: fuzzy set, intuitionistic fuzzy set, fuzzy event, intuitionistic fuzzy event, fuzzy probability, intuitionistic fuzzy probability.

1 INTRODUCTION

Probability theory has been a powerful tool for dealing with uncertainty. For instance, it has been able to handle uncertain events exemplified by the temperature of air tomorrow. Assuming the availability of some past data, one could has made it possible to calculate the probability of, say, the air temperature between 20° C and 25° C . However, events have had to be precisely defined.

In practice, however, people often use events which are imprecisely specified, and questions like "what is the probability of a nice weather tomorrow" are considered well valid in human discourse, though they are beyond the scope of classic probability theory.

Since the very beginning of fuzzy sets theory it has been clear that one of key issues would be to be able to devise some formal tools for expressing probabilities of such imprecisely specified events.

And, indeed, it was Zadeh (1968) who first introduced the concept of a fuzzy event, and introduced the concept of its probability by a natural extension of the probability of a nonfuzzy event. In case the probability was specified as a real number from the unit interval. Later on, some extensions have been proposed whose essence was the introduction of a fuzzy probability of a fuzzy event meant as a fuzzy number in the unit interval.

In the mid-1980s Atanassov introduced the concept of an intuitionistic fuzzy set [cf. Atanassov (1986)]. Basically, his idea was that unlike in conventional fuzzy sets in which imprecision is just modelled by the membership degree from [0,1], and for which the nonmembership degree is just automatically the complementation to 1 of the membership degree, in an intuitionistic fuzzy set both the membership and nonmembership degrees are numbers from [0,1], but their sum is not 1. Thus, one can express a well known psychological fact that a human being who expresses the degree of membership of an element in a fuzzy set, very often does not expresses, when asked, the degree of nonmebership as the complementation to 1.

This idea has led to an interesting theory whose point of departure is such a concept of an intuitionistic fuzzy set.

In this article we introduce the concept of an intuitionistic fuzzy event and a probability of an intuitionistic fuzzy event. The proposed probability lies in an interval and is such that when considered intuitionistic fuzzy event becomes a fuzzy event (by making the hesitancy margin equal to zero) the proposed interval is reduced to the probability of a fuzzy event defined by Zadeh.

Next, we replace an intuitionistic fuzzy event by

the best corresponding fuzzy event [cf. Szmidt and Kacprzyk 1998b]. In effect the probability of such an event lies in the middle of the previously obtained interval and is consistent with the solution proposed by Gesternkorn and Manko (1990) and Zadeh.

Next, we consider some other properties of the interval in which a probability of an intuitionistic event lies.

2 MAIN DEFINITIONS AND RESULTS

As opposed to a fuzzy set (Zadeh, 1965) in $X = x$, given by:

$$A' = \{< x, \mu_{A'}(x) > | x \in X\} \tag{1}$$

where $\mu_{A'} : X \rightarrow [0,1]$ is the membership function of the fuzzy set A' , an intuitionistic fuzzy set (Atanassov, 1986) $A \in X$ is given by

$$A = \{< x, \mu_A(x), \nu_A(x) > | x \in X\} \tag{2}$$

where: $\mu_A : X \rightarrow [0,1]$ and $\nu_A : X \rightarrow [0,1]$ with the condition

$$0 \leq \mu_A(x) + \nu_A(x) \leq 1 \tag{3}$$

and the numbers $\mu_A(x)$, $\nu_A(x) \in [0,1]$, denote the degree of membership and non-membership of x to A respectively.

For each intuitionistic fuzzy set in X, we will call

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x) \tag{4}$$

the *intuitionistic index* of x in A and it is a hesitancy degree of whether x belongs to A [cf. Atanassov 1986, 1989, 1994a, 1994b]. It is obvious, that $0 \leq \pi_A(x) \leq 1$ *for each* $x \in X$.

Intuitionistic fuzzy sets can be useful in modeling many real situations, and, examples in negotiation processes are given in Szmidt and Kacprzyk (1996a, 1996b, 1996c, 1997a, 1998f, 1998g]).

Zadeh (1968) published the first work dealing with a probability of fuzzy events. By a fuzzy event A' he meant a fuzzy subset of the elementary space E' , i.e. $A' \subset E'$ whose membership function $\mu_{A'}(x)$ was Borel measurable.

Definition 1. Let us assign to every element of a fuzzy event $A' \subset E' = \{x_1, \ldots, x_n\}$ its probability of occurrence, i.e. $p(x_1), \ldots, p(x_n)$. Probability of a fuzzy event is equal to

$$p(A') = \sum_{i=1}^{n} p(x_i)\mu_{A'}(x_i) \tag{5}$$

Definition 2. By an intuitionistic fuzzy event A we will mean an intuitionistic fuzzy subset belonging to the elementary event space E , i.e. $A \subset E$ whose membership function $\mu_A(x)$ and hesitancy margin $\pi_A(x)$ are Borel mesurable.

Definition 3. Let us assign to every element of an intuitionistic fuzzy event $A \subset E = \{x_1, \ldots, x_n\}$ its probability of occurence, i.e. $p(x_1), \ldots, p(x_n)$.

Minimal probability $p_{min}(A)$ of an intuitionistic fuzzy event A is equal to

$$p_{min}(A) = \sum_{i=1}^{n} p(x_i)\mu_A(x_i) \tag{6}$$

Maximal probability $p_{max}(A)$ of an intuitionistic fuzzy event A is equal to

$$p_{max}(A) = p_{min}(A) + \sum_{i=1}^{n} p(x_i)\pi_A(x_i) \tag{7}$$

so probability of an event A is a number from the interval $[p_{min}(A), p_{max}(A)]$, or

$$\begin{aligned} p(A) \quad \in \quad & [\sum_{i=1}^{n} p(x_i)\mu_A(x_i), \sum_{i=1}^{n} p(x_i)\mu_A(x_i) + \\ & + \sum_{i=1}^{n} p(x_i)\pi_A(x_i)] \end{aligned} \tag{8}$$

Lover boundary of the interval (8) i.e., minimal probability $p_{min}(A)$ (6) gives "sure" or quaranteed probability that an event A will occur.

Upper boundary - maximal probability $p_{max}(A)$ (7) gives the highest possible probability that an event A will occur. The highest probability is achieved when hesitancy margins support occurence of an event A.

The difference between the upper and the lower boundary of the interval (8) i.e., $p_{max}(A) - p_{min}(A)$ is the unsureness of occurence of an intuitionistic event A, i.e.:

$$p_{max}(A) - p_{min}(A) \quad = \quad \sum_{i=1}^{n} p(x_i)\pi_A(x_i) \tag{9}$$

Certainly, the intervals of the estimated probabilities are not totally arbitrary.

- If $p_{max}(A) \leq 1$, then the estimated probabilities are in the interval

$$[0, p_{max}(A)] \tag{10}$$

- If $p_{min}(A) \geq 0$, then the estimated probabilities are in the interval

$$[p_{min}(A), 1] \tag{11}$$

Let us notice that the proposed probability (8) reduces to Zadeh's probability measure (5) in the case when the considered intuitionistic fuzzy event is replaced by ordinary fuzzy set (when assuming that the hesitancy margin is equal to zero). But Atanassov (1994a) proposed a general method of transformation of intuitionistic fuzzy sets into fuzzy sets. It can be done by the following operator

$$\begin{aligned} D_\alpha(A) &= \{< x,\ \mu_A(x) + \alpha\pi_A(x),\ \nu_A(x) + \\ &+ (1-\alpha)\pi_A(x) > | x \in X\} \end{aligned} \tag{12}$$

where $\alpha \in [0,1]$

Anyway, the definition of $D_\alpha(A)$ does not tell the differences among sets obtained from different parameters α. In Szmidt and Kacprzyk (1998b) we have showed that a good approximation of an intuitionistic fuzzy set via a fuzzy set is given by the operator (12) with α equal to 0.5, i.e.

$$\begin{aligned} D_{0.5}(A) &= \{< x,\ \mu_A(x) + 0.5\pi_A(x),\ \nu_A(x) + \\ &+ 0.5\pi_A(x) > | x \in X\} \end{aligned} \tag{13}$$

Having the above in mind, and taking into account Zadeh's approach (5) we can calculate a probability of such a fuzzy set corresponding to an intuitionistic fuzzy set. The probability of such a fuzzy set lies in the middle of the interval (8), i.e.

$$\begin{aligned} p_{0.5}(A) &= \sum_{i=1}^{n} p_i(x_x)[\mu_A(x_i) + 0.5\pi_A(x_i)] = \\ &= 0.5[p_{min}(A) + p_{max}(A)] \end{aligned} \tag{14}$$

It is worth noticing that (14) can be stated as

$$\begin{aligned} p_{0.5}(A) &= \sum_{i=1}^{n} \frac{p(x_i)}{2} \cdot [2\mu_A(x_i) + \pi_A(x_i)] = \\ &= \sum_{i=1}^{n} \frac{p(x_i)}{2} \cdot [\mu_A(x_i) + 1 - \nu_A(x_i)] \end{aligned} \tag{15}$$

The last expression was proposed in Gerstenkorn and Mako (1990) as a probability of an intuitionistic fuzzy event. It was also demonstrated that $p_{0.5}(A)$ satisfies the properties from the Kolmogorov axiomatic definition of probability as well as other properties characteristic of the classical probability.

Next, when $\pi_A(x_i) = 0$, i.e. for a classical fuzzy set, (14) obviously is reduced to probability of a fuzzy set in the sense of Zadeh's probability.

Let us look at some properties of the border probabilities of A and $\overline{A}$, where $\overline{A}$ denotes the complement of A defined (Atanassov, 1989) the following way

$$\overline{A} = \{< x, \nu_A(x),\ \mu_A(x), > / x \in X\} \tag{16}$$

So we can assign probability that an opposite event $\overline{A}$ will occur (or that an event A will not occur).

Definition 4. Let us assign to every element belonging to the complement of an intuitionistic fuzzy event A i.e., to $\overline{A}$ its probability of occurence, i.e. $p(x_1), \ldots, p(x_n)$.

Minimal probability that an event $\overline{A}$ will occur is equal to

$$p_{min}(\overline{A}) = \sum_{i=1}^{n} p(x_i)\nu_A(x_i) \tag{17}$$

Maximal probability that an event $\overline{A}$ will occur is equal to

$$p_{max}(\overline{A}) = \sum_{i=1}^{n} p(x_i)\nu_A(x_i) + \sum_{i=1}^{n} p(x_i)\pi_A(x_i) \tag{18}$$

Having in mind that $\mu_A(x_i) + \nu_A(x_i) + \pi_A(x_i) = 1$ we have

$$\begin{aligned} p_{min}(\overline{A}) &= \sum_{i=1}^{n} p(x_i)\nu_A(x_i) = \\ &= \sum_{i=1}^{n} p(x_i)[1 - \mu_A(x_i) - \pi_A(x_i)] = \\ &= \sum_{i=1}^{n} [p(x_i) - p(x_i)\mu_A(x_i) - p(x_i)\pi_A(x_i)] = \\ &= 1 - p_{max}(A) \end{aligned} \tag{19}$$

i.e., minimal probability that an opposite event $\overline{A}$ will occur is equal to one minus maximal probability that an event A will occur.

Next,

$$\begin{aligned} p_{max}(\overline{A}) &= \sum_{i=1}^{n} p(x_i)[\nu_A(x_i) + \pi_A(x_i)] = \\ &= \sum_{i=1}^{n} p(x_i)[1 - \mu_A(x_i)] = \\ &= \sum_{i=1}^{n} p(x_i) - \sum_{i=1}^{n} p(x_i)\mu_A(x_i) = \\ &= 1 - p_{min}(A) \end{aligned} \tag{20}$$

what means that maximal probability of an opposite event $\overline{A}$ will occur is equal to one minus minimal probability of an event A.

Next, from (19) we have

$$p_{min}(A) + p_{min}(\overline{A}) = p_{min}(A) + (1 - p_{max}(A)) = \\ = 1 - \sum_{i=1}^{n} p(x_i)\pi_A(x_i) \leq 1 \quad (21)$$

and from (20)

$$p_{max}(A) + p_{max}(\overline{A}) = p_{max}(A) + 1 - p_{min}(A) = \\ = 1 + \sum_{i=1}^{n} p(x_i)\pi_A(x_i) \geq 1 \quad (22)$$

3 FINAL REMARKS

We proposed the probability measure for intuitionistic fuzzy events which is consistent with Zadeh's approach for fuzzy sets.

Next, we investigated some properties of intervals in which the probabilities of an intuitionistic fuzzy event and its complement lie.

The proposed formulas handle two types of imperfection of information - uncertainty (randomness)expressed by probability theory on the one hand, and imprecision (fuzziness) expressed in this article by intuitionistic fuzzy set theory. Both types of imperfection play central role in decision making.

References

[1] K. Atanassov , Intuitionistic fuzzy sets, Fuzzy Sets and Systems, 20 (1986) '87-96.

[2] K. Atanassov, More on intuitionistic fuzzy sets, Fuzzy Sets and Systems 33 (1989) 37-46.

[3] K. Atanassov, New operations defined over the intuitionistic fuzzy sets, Fuzzy Sets and Systems 61 (1994a) 137-142.

[4] K. Atanassov, Operators over interval valued intuitionistic fuzzy sets, Fuzzy Sets and Systems 64 (1994b) 159-174.

[5] K. Atanassov , Intuitionistic Fuzzy Sets: Theory and Applications (to appear).

[6] T. Gerstenkorn and J. Mako, Probability of fuzzy intuitionistic sets, BUSEFAL, Vol.45, (1990), 128-136.

[7] E. Szmidt and J. Kacprzyk, Intuitionistic fuzzy sets in group decision making, Notes on IFS 2 (1996a) 15-32.

[8] E. Szmidt and J. Kacprzyk, Group decision making via intuitionistic fuzzy sets, Proceeding of FUBEST'96 (Sofia, Bulgaria, 1996b) 107-112.

[9] E. Szmidt and J. Kacprzyk, Remarks on some applications of intuitionistic fuzzy sets in decision making, Notes on IFS 2 (1996c) 22-31.

[10] E. Szmidt and J. Kacprzyk, Intuitionistic fuzzy sets for more realistic group decision making, Proceedings of TRANSITIONS'97 (Warsaw, Poland, 1997a) 430-433.

[11] E. Szmidt and J. Kacprzyk, On measuring distances between intuitionistic fuzzy sets, Notes on IFS 3 (1997b) 1-13.

[12] Szmidt E. and Kacprzyk J. (1998a) Intuitionistic Fuzzy Linguistic Quantifiers. Notes on IFS, Vol.3, No.5. pp. 111-122.

[13] Szmidt E. and Kacprzyk J. (1998b) A Fuzzy Set Corresponding to an Intuitionistic Fuzzy Set. Int. Journal of Uncertainty and Knowledge Based Systems, Vol.6, No.5, pp.427-435.

[14] Szmidt E. and Kacprzyk J. (1998c) Intuitionistic Fuzzy Set Theory and Mass Assignment Theory: Some relations. Notes on IFS, Vol.4, No.1, pp. 1-7.

[15] Szmidt E. and Kacprzyk J. (1998d) Entropy for Intuitionistic Fuzzy Sets. Accepted for publication in Fuzzy Sets and Systems.

[16] Szmidt E. and Kacprzyk J. (1998e) Distances Between Intuitionistic Fuzzy Sets. Accepted for publication in Fuzzy Sets and Systems.

[17] Szmidt E. and Kacprzyk J. (1998f) Group Decision Making under Intuitionistic Fuzzy Preference Relations. Proceedings of the 7th Int. Conference IPMU'98 (Paris, La Sorbonne, July 6-10), pp. 172-178.

[18] Szmidt E. and Kacprzyk J. (1998g) Applications of Intuitionistic Fuzzy Sets in Decision Making. Proceedings of the 8th Congreso EUSFLAT'9 (Pamplona, Univ. De Navarra, September 8-10), pp. 150-158.

[19] L.A. Zadeh (1965) Fuzzy sets. Information and Control, 8 338-353.

[20] L.A. Zadeh (1968) Probability measures of fuzzy events. J.Math.Anal.Appl. Vol.23, pp..421-427.

[21] L.A. Zadeh (1983a) A computational approach to fuzzy quantifiers in natural languages. Comput. Math. Appl., Vol. 9, No. 1, pp. 149-184.

[22] L.A. Zadeh (1983b) The role of fuzzy logic in the management of uncertainty in expert systems. Fuzzy Sets and Systems, Vol. 11, pp. 199-227.

THE OPTIMIZATION OF CAPACITY UTILIZATION WITH A FUZZY DECISION SUPPORT MODEL

Rudolf Felix

FLS Fuzzy Logik Systeme GmbH
Joseph-von-Fraunhofer-Str. 20
44227 Dortmund, Germany
phone: (+49) 231 / 97 00 921 fax: (+49) 231 / 97 00 929
e-mail: felix@fuzzy.de

Summary

The assembly of cars is a production process, whose regulation must respond to the disturbances and fluctuations that often arise in such processes. Assembly is carried out at stations connected in series, where specialized teams of workers are employed. Depending on the model variant, each model that passes an assembly area creates a different workload. The order of transport of models into the assembly area is very important for the degree of utilization of capacity at each assembly station. The problem is to fix the order in such a way that the utilization of capacity is as constant as possible and as near to 100 % as possible, in order to guarantee low-cost assembly for the given long-term production plans. This paper presents an industrial application of a fuzzy decision support model for the supervision of capacity optimization in the car assembly process.

Keywords: fuzzy decision making, relationships between goals, capacity utilization, production planning.

1 INTRODUCTION

The assembly of cars is a production process, whose regulation must respond to the disturbances and fluctuations that often arise in such processes. The blocking and release of different models and special equipment require flexible reactions by the staff supervising the control system, in order to guarantee that utilization of the assembling capacity is as regular as possible. To achieve this, the supervisor often adjusts the production process manually by modifying the order of introduction of the car bodies into the assembly area and the mixture of special equipment required. Due to the complexity of the effects of different mixes of equipment on the utilization of capacity, adequate manual supervision of the assembly process is very difficult. If all restrictions are to be incorporated into the supervisor's decision, an automated decision system is required.

2 THE OPTIMIZATION PROBLEM

The complexity of the optimization problem to be solved can be described as follows. The production of the vehicles is performed in various successive steps. First, the car bodies are preformed. Then, these preformed car bodies are painted and stored in the so-called high store.

According to the production program, the painted car bodies are transported into the assembly area, where modules such as undercarriage or motor, or special equipment such as electric window lifters, sunroofs or air-conditioning systems, depending on the model variant, are assembled.

The assembly is carried out at various assembly stations connected in series, where specialized teams of workers are employed. Depending on the model variant, each model that passes the assembly area creates a different workload at the individual assembly stations.

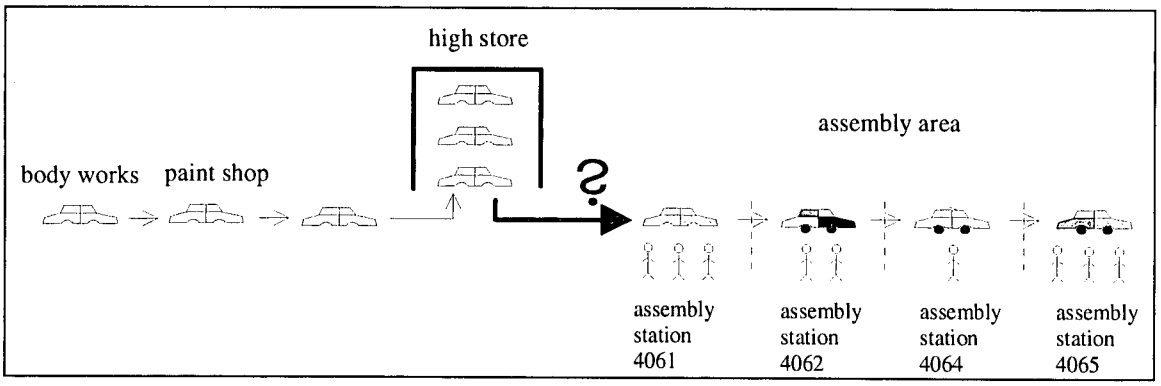


Figure 1: Assembly process of cars

The order of transport of the models into the assembly area is very important for the degree of utilization of capacity at the assembly stations. The problem is to fix the order in such a way that the utilization of capacity is as constant as possible and as near to 100 % as possible, in order to guarantee low-cost assembly with the given long-term production plans. Figure 1 gives an overview of the structure of the assembly process. Figure 2 shows a practical capacity utilization diagram for a set of different assembly stations.

Factors such as the blocking or release of model variants, special equipment, colours, etc., in the areas of preforming and painting alter the number of car bodies to be optimized in the high store. That is why it is necessary for these factors to influence the optimization parameters,

in order to determine the sequence in which the vehicles are to be taken out of the high store. The optimization parameters indicate how many items of every model variant and of special equipment (for example model No. 3 without sunroof) are to be assembled, expressed in number of pieces for the relevant day.

The complexity of the effects of blocking on the number of pieces of model variants and special equipment available renders manual supervision nearly impossible, if all restrictions concerning the assembly and the utilization of capacity are to be met.

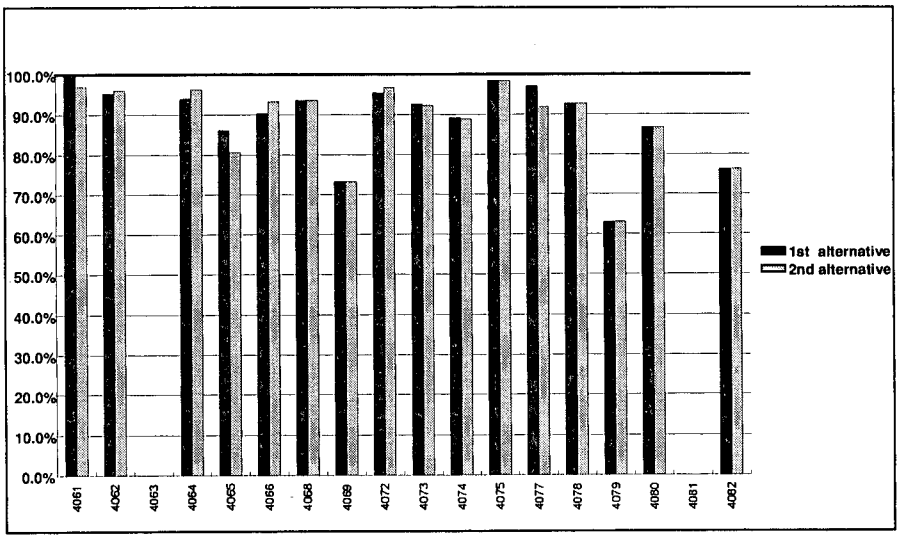


Figure 2: Capacity utilization of assembly stations 4061 - 4082

The problem of optimal capacity utilization of the assembly stations described above can be understood as a multicriterial optimization problem in the following way. Every assembly station's degree of utilization is viewed as an optimization goal. This means that if we have, for instance, n assembly stations a_i with the corresponding indices $i = 1,..., n$, then there is a vector of goals (g_{a1}, ... , g_{an}) to be maximized. The goal function g_{ai} has values belonging to the interval [0,1] where 0 stands for 0 % utilization of capacity of the assembly station g_{ai} and 1 stands for 100 % utilization of that station. For simplicity, we write g_i instead of g_{ai}. The domain of each goal

function is a vector of quantities of equipment (e_{j1}, ... , e_{jm}) to be assembled into the car bodies in the planning period, and m is the number of types of equipment. Each e_{jk} means the quantity of equipment items of the type j. For simplicity again, we write e_j instead of e_{jk}. Every vector (e_1, ... ,e_n) is a possible decision alternative. The goal function g_i expressing the capacity utilization of the assembly station a_i can be understood as

$$g_i : (e_1, \ldots, e_n) \longrightarrow [0,1] , \; i \in \{1, \ldots, n\} . \qquad (1)$$

For each decision alternative, for example, the vector (e_1, ... ,e_n) its impact on the capacity utilization of a_i determines the corresponding value of g_i. However, since the impacts cannot be described in an analytical way, an estimate has to be calculated for each vector (e_1,....., e_m). The estimates are calculated by an analysis algorithm based on statistical data that reflect the distribution of degrees of different capacity utilization corresponding to different equipment vectors. The statistical distribution was used to estimate the desired impacts and thus to derive an estimation algorithm providing for the calculation of values of every possible (e_1, ... ,e_n)[1].

In conclusion, the optimization problem to be treated can be defined as

$$\textbf{Max } (g_1, \ldots, g_n) \qquad (2)$$

Note that every g_i is a fuzzy goal function over the space of the decision alternatives (e_1, ... ,e_n).

There are two more important aspects to consider.

1. The assembly stations represented as capacity utilization functions are not always independent of each other. Moreover, they may affect each other in the sense that a better capacity utilization of a particular assembly station affected by an e_k may cause a worse utilization of another station and, vice versa, the goal functions $g_i , i \in \{1, \ldots, n\}$ interact. This means that an adequate decision model for the solution of the capacity utilization problem should be able to deal adequately with interacting goals. Relationships between goals such as co-operation and competition must be recognized and taken into account.

2. Since different assembly stations may cause more or less significant costs, it should be possible to attach varying priorities to the different goals g_i.

[1]Because of the limited number of pages the estimation algorithm is not disscused in this paper.

3 INDUSTRIAL APPLICATION

The optimization problem described in Section 2 has been solved using the software tool FuzzyDecisionDesk [2], which is based on the decision model discussed in [1]. The System works at BMW, Factory Regensburg on an IBM compatible PC with an Intel 486 DX2/66 MHZ processor and a graphical user interface under Microsoft Windows.

4 CONCLUSIONS

The application presented in this paper is based on the concept of fuzzy relationships between goals. It shows that the decision model helped to solve a complex optimization problem in the field of capacity utilization in the assembly process of an automotive company. Because of the relatively high problem- independence of the decision model, there is strong evidence that the decision model can be applied in a variety of fields.

5 REFERENCES

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

[2] **FLS Fuzzy Logik Systeme GmbH**
FuzzyDecisionDesk, User Manual for Version 1.7

VIRTUAL REALITY SYSTEM "ROBOT-SECRETARY": METHODOLOGICAL PROBLEMS AND NEURO-FUZZY APROACH

David Shapiro

Institute for Problems of Computer Technology and Informatization,
113114 2 Kozhevnicheskii per.4/6,
Moscow, Russia, Tel.(095) 235-71-66, Fax (095) 235-52-67

Summary

The methodological problems of the processing of imagery knowledge about the world on the brain-like specifics basis and corresponding formal presentations for the virtual reality system "Robot-Secretary" are presented.

Keywords: virtual reality, cerebral asymmetry, neuro-fuzzy algorithms, perception, understanding, formation of the result.

1 INTRODUCTION

The present time is characterized by the greate acivements of the computer technologies (structures, memory, frequency and methods of the calculations) and wide differentiation of the neurochips. One priority of the future stage of the civilization will be "the Personal-orientness" for the diffrent forms the User's activity (business, culture, medicine etc). These possibilities allow to realize the important achivements of the neuropsycholinguistical investigations for the both approaches:

- the simulation of the neuropsycholinguistical processes of the processing of knowledge about the world (Perception [P]—Understanding [U]—Formation of the resilt [F]);

- the formation of the procedures of the (P-U-F) for the personal-oriented systems of the virtual reality (VRS).

The applied tasks are characterized by the following aspects:

- wide differentiation of the forms of the perception of knowledge about the world(sensorical channels, forms of the mematext, types of the bearers of the information),

- preferable forms of the presentation of imagery knowledge,

- important influence of the User's specifics for each prosess,

- development of the User's cognitive ability and creative activity.

These aspects are embraced by means of the technologies of the VRS.

2 METHODOLOGICAL PROBLEMS

2.1 MULTISIGNIFICANT IMAGES IN DESCRIPTION OF WORLD

The multisignificant images are characterized by the both:

- the complexity presentation of the Gestalt and unsynonymous expression;

- the multisensotical (multimodal) presentation (audio, visual, tactile, small) by means of the different forms of the metatext.

The realization of the (P-U-F) stages is based on the three methodological conceptions: "Ego in the World", "Meaning—Text—Meaning" and "Two Forms of Activity".

The central problem of the U-stage is the discovery and interpretation of the Meaning (DIM) which is contained in the concrete form of the metatext [MET]. The [MET] is presented by the different sensorical channels (audio, video, tactile, smell) and different types of the bearers (printed text-order, poetic, art text, scientific text, notes, hierogliphs; visual image-expression of the faces and poses, graphics, structures, picture on the screen, TV-presentations ; audio text- music play, intonational

chatracteristics of the speech, natural sounds, technical sounds; "smell accord"). The discrimination of this forms of the metatexts is characterized by the type of the channel and , influence of the art, science, life or nature, influence of the temporal factors (Past-Present-Future) and degree of the temporal presentations (at To or at the interval).

The meaning of the concrete realization of the metatext is Sc (which is the membership to the set of the meanings Sc E {S}).

The DIM-problems are characterized by the determination of the interconnections [MET] $\Longleftrightarrow$ Sc.

The types of the DIM -problems:

Type-I: detemination of the degree of membership of the concrete meaning to the interval

$$[\mathrm{MET}] \Longrightarrow \{\mathrm{M(Sc)} \longrightarrow [\mathrm{S1}, ..., \mathrm{Sn}]\mathrm{vr}\},$$

where: Mis the membership function, Sc is the concrete sign of the meanig, v is the type of the sensotical channel, r is the concrete feature.

Type -II: determination of the degree of the membership of the multisignificant image to the certain meanings (when the information is presented by the different sensorical channels, or characterizes the complex context)

$$[MET] \Longrightarrow \begin{matrix} \{M'(Sc)for[, ..., S, ...]'\} \\ \{.....\} \\ \{M''(Sc)for[, ..., S, ...]''\}. \end{matrix}$$

Type -III: creating of the new image for the concrete sign of the meaning (membership function)

$$\{\mathrm{M(Sc)vr}\} \Longrightarrow [\mathrm{MET}]\mathrm{vr}$$

The examples of the realization for the Type-I DIM-problems are interpretation of the expressions of the concrete faces (more kind, then jolly) for the Type-II DIM-problems are interpretation of the situation "Distruction of the Titanik".

The formation of the new image on the procedure "the hysteresis of the perception" basis is the samples example of the Type-III DIM-problem. These are only the illustrative examples. The real tasks need the very complex Base of Knowledge and very labour intensive works for the exposure of imagery knowledge in the concrete classe of the context and environment. The formal procedures are based on the dual and multiconnections neural models and neuro-fuzzy algorithms.

2.2 HUMAN'S SPECIFICS FOR PROCESSING OF IMAGERY KNOWLEDGE

Knowledge about the world are presented by means of the different forms of the [MET] and are processed on the Human's specifics basis (discovery of the meaning and formation of the new metatext). Such procedures are based on the Human's possibilities cognitive ability and creative activity which are determined by the highest brain's functions and structural-functional specifics.

The highest brain's functions include the following: attention, cerebral mechanisms of the processong of imagery knowledge, learning, imagery memory, assotiativeness which are based on the interconnected structural-functional features: dualism, plasticity,adaptiveness, multilayerness, multiconnectionness, self-organization).

The brain's Dualism is characterized by the specifics of the cerebral asymmetry mechnisms (parallele different -functional Left and Right hemispherical procedures). The brain's Placticity is realized on the both (Churchland, 1989): neural layer (neural interconnections) and on the "learning-behavioral" layer which is characterized by the cognition and learning of the influence of the environment and personal sensation.

The brain's adaptiveness is characterized by the ability of the dynamic change of the characteristics for the better processing of the input pattern. The adaptive procedures are used two types procedures: dynamic change of the structures and dynamic change of the axiomatic basis.

The concrete procedures of the dynamic change of the axiomatic basis are realized on the (P-U-F) stages. The realization of the P-stage is characterized by the choising of the suitable method (scanning on the screen, movement on thr contour, segmentation, whole embrace of the Gestalt) for the concrete input pattern, Human's state and environment.

The realization of the U-stage is characterized by the estimatiion of the elements and interconnections, logical-

combinatory and analitical procedures; whole estimation of the Gestalt, interpretation of the meaning and associations. The realization of the F-stage is characterized by the choising of the suitable form of the reaction (verbal, operational and extralinguistical).

3 FUZZY AND NEURO-FUZZY ALGORITHMICAL APPROACH

3.1 FUZZY CATEGORIES AND MEANING

The types of the functional characteristics for the presentation of the changing of the meanings by four types are determined:

Type-I: Monotonic changing;

Type-II: Unmonotonic changing;

Type-III: Central point of the several meanings;

Type-IV: Double meaning.

The qualitative scales are characterized by the "supporting elements"(SE), (for example, Normal, High, Low - signs) and the formation of the another elements in the internal interval (the interpolational procedure) and external interval (the extrapolational procedure).

The qualitative scales ("beauty", "not beauty", "not enough beauty") and the SE are determined on the experience basis. This experience is formed on the genotype and phenotype layers ("mother's voice") and behavioral procedures basis.

3.2 U-STAGE: SPECIFICS OF THE LHP AND RHP

The LH-procedures realize the following neuro-informational mechanisms: formation of the qualitative scales $\{q\}$, estimation of the features of the world, logical-combinatory, probability and analytical procedures, decision making, analysis of the grammatical constructions and interpretation of the strong associations ("vodka-butl"). The formal LHP procedures are based on the using of the fuzzy categories and estimations $\{q\} : X \longrightarrow Y$, fuzzy algorithms and fuzzy production rules. The fuzzy categories are presented by means of the membership function M (X) and fuzzy decision making by the membership surface Re, $M(Re) = O\ (Y1, ..., Yn)$, (where Re is the rational recommendation, Yi is the estimated parameters) and the procedures "IF [< . >,...,< . >] THEN < . >" and also by the fuzzy graphs for the estimation of the structures.

This procedures are formed on the human-like learnig basis. The formation of the qualitative multielement scales is realized on the human's professional and personal experience basis.

The LH-procedures present the estimated characteristics of the world by $Y = (Y1, ..., Yn)$.

The RH-procedures realize the the whole estimation of the Gestalt, interpretation of the images and wide associations, discovery of the meaning.

The RH procedures form the the personal oriented model of the world. These procedures realize the answere on the question 1.

The interior implicite model (IIM) forms the personal presentation about characteristics of the world, their estimated qualitative shades ("mother voice", "beauty", "very much" etc) and personal option ("common sense is own option"). The IIM is the specialized fragment of Base Knowledge. on the experience basis by means of the set $\{Y1m, ..., Ynm = Ym\}$.

The characteristics of the Gestalt are determined by the whole estimation on the qualitative scalel basis

$$G\{X, f(X)\} \Longrightarrow \{< Yg >\}$$

The problem of the global estimation of the characteristics of the world and exposure of the meaning is determined by the following interconnectons: $\{< Yg >, < Ym >, < Y >\}$.

The formation of the result (F) is the LH procedure which is characterized by the following stages:

- determination of the specifics of the input signals,
- determination of the specifics of the interior presentations about the human's state and environment,
- selection of the forms for the concrete behavioral realization (verbal, operational and extralinguistical).

3.3 RHP (CHORUS)

The formation of the brain-like procedures of the understanding needs of the using of all power of the brain's

structural-functional specifics and features of the high functions. This is a new methodology which is based on the ideas of M.L. Tsetlin (Tsetlin's Models, TM.; 1969). Three principles characterize of the TM: "Polyfunctionalism of the neural pools"; "Priciple of the shortest interconnection"; "Games of the automats in the stochastic space". The TM allows to aggregate the different brain's features: the regions, structures of functioning and local behavioral procedures, influence of the input signals and subject's specifics, local games which are characterized by the "intensity" of the estimations and the local strategy.

CONCLUSION

The applications contain the interpretation of the expressions of the faces, poses and intonational characteristics of the speech, articulations for the inner procedures (IP) of the system "Robot-Secretary" for the different concrete problems: "Multisensorical personal discourse", "Attendance of museums and libraries for human with sensorical and operational violations" (Shapiro, 1996), "Walks on Rinascimento Roma with Stendal" (Shapiro, 1998), simulation of the neuropsycholinguistical processes (interpretation of the metaphors, expression of the faces and poses, and intonational characteristics of the speech) for the formation of the discourse in the systems of countrylogy.

References

[1] P.Churchland (1989). Neurophilosophy. Badford Book, MIT Press, London.

[2] D.I.Shapiro (1996). Virtual Reality Systems: Museums and Libraries, Proc. of EVA'96, London, pp (12-1)−12-5.

[3] D.I.Shapiro (1997), Dual Processing of Multisensorical Pattern for Virtual Reality: Neuro-Fuzzy Approach, Proc. of EUFIT'97, Aachen, pp 1983-1987.

[4] D.Shapiro (1998). Virual Walks on Rinascimento Roma with Stendal, Proc. of EVA98, Florance, pp 25-31.

[5] R.R.Yager (1994). Modeling and Formulating Fuzzy Knowledge Base Using Neural Networks, N.N.7, N 8, pp 1273-1284.

GENERALIZED NEURAL NETWORKS FOR FUZZY MODELING

K.-H. Temme, R. Heider and C. Moraga

Department of Computer Science, Computer Engineering and Computing Education
University of Dortmund
44221 Germany

{temme | heider | moraga}@LS1.informatik.uni-dortmund.de

Summary

Neuro-fuzzy modeling has been intensively studied since the early nineties. Recently a method has been disclosed, that uses a classical feedforward neural network with just one hidden layer. Nodes of the hidden layer use the logistic function as activation function meanwhile the output node has a linear activation function. This paper introduces a generalization of the logistic function and evaluates its capabilities with respect to neuro-fuzzy modeling. It is shown that a product-generated symmetric summation provides an exact interpretation of the activity of each node of the hidden layer.
Keywords: Neuro-fuzzy-modeling, aggregation

1. INTRODUCTION

Fuzzy modeling based on neural networks is a well established area within Computational Intelligence. Most work has been done using RBF nets [16] or special feedforward networks, where the first hidden layers realize the fuzzification (association of physical crisp inputs to linguistic terms) and compute the corresponding degree of satisfaction. Further layers compute the fuzzy inference and, if needed, additional layers will cope with defuzzification. Probably one of the best known representatives of this class is the ANFIS system [13], since it is presently included in Matlab, however similar networks have been discussed in the literature (see *e.g.* [6], [8], [11], [18], [19], [20]).

A recent contribution [1] has opened very attractive possibilites for effectively applying the simplest classical feedforward neural network for fuzzy modeling. The neural network has only one hidden layer where the processing nodes have a logistic activation function $f(wsi) = (1+\exp(-wsi))^{-1}$, ("wsi" denotes here "weighted sum of inputs"), and a linear output node. A neural network of this kind satisfies the conditions for universal approximation [9], [15] which is roughly equivalent to saying that for any physical system there exists at least one neural network of this kind, which can identify the system with arbitrary precision if enough representative performance examples are available. Furthermore it has been shown [1] that these neural networks, under an especial interpretation (see below) are equivalent to additive fuzzy systems [13].

The present paper introduces a generalization of these neural networks however preserving their interpretability as additive fuzzy systems.

The rest of the paper is organized as follows. In the next section an extended class of activation functions will be discussed and their characteristic properties will be explained. Furthermore it will be shown, that the method introduced in [1] may be applied to the new neural networks making them appropriate for fuzzy modeling. An example followed by a summary and conclusions will close the paper.

2. GENERALIZED LOGISTIC FUNCTIONS

Definition 1:

A generalized logistic function $f: R \rightarrow [0,1]$ is given by:

$$f(x) = \frac{1}{1+k^{-x}} \qquad (1)$$

where $k \in R$ and $k > 1$.
It becomes apparent that the classical logistic function is a particular case of (1), when $k = e$. Equation (1)

represents a family of "soft step" functions whose steepness is controlled by the parameter k.

Observations:

- f(x) is overall differentiable with respect to x, *i.e.* the generalized logistic function may be used in neural networks trained with the backpropagation algorithm
- Equation (1) satisfies the conditions of continuity, $lim_{x\to-\infty}f(x) = 0$ and $lim_{x\to+\infty}f(x) = 1$, *i.e.* f(x) is a *squashing* function [9]. It follows that neural networks with one hidden layer using the generalized logistic function as activation function and an output linear node are universal approximators [9].
- Equation (1) is invertible. It follows that neural networks using *allover* the generalized logistic function as activation functions are universal approximators [15].
- The partial derivative of f(x) with respect to k is continuous and well defined, *i.e.* it is possible to extend the backpropagation algorithm to also train *the best k* at every node of the hidden layer. Similar work done to optimize the steepness of the classical logistic function showed that a faster learning and a better generalization could be achieved [7].

3. A FUZZY LOGIC INTERPRETATION

In [1] it was proven that the activity of a node of a neural network using a sigmoide activation function (k = *e* in the notation of the present paper) may be interpreted as a fuzzy *if-then* rule, if a special aggregation of the premises is used. In what follows it will be shown, that the same interpretation may be extended to the case of the generalized logistic function.

Lemma 1:

Let f be a generalized logistic function. There exists a bijection Ψ: (R, +) $\rightarrow$ ((0,1), *) producing the following transformation:

$$f(x+y) = f(x) * f(y) \qquad (2)$$

where the operation "*" will be expressed explicitly below (see eq. (3)).

Proof:

$$f(x+y) = \frac{1}{1+k^{-x-y}} = \frac{1}{1+k^{-x}k^{-y}}$$

but since

$$k^{-x} = \frac{1-f(x)}{f(x)} \quad \text{and} \quad k^{-y} = \frac{1-f(y)}{f(y)} \quad \text{then}$$

$$f(x+y) = \frac{1}{1+\left(\frac{1-f(x)}{f(x)}\right)\left(\frac{1-f(y)}{f(y)}\right)}$$

$$f(x+y) = \frac{f(x)f(y)}{(1-f(x))(1-f(y))+f(x)f(y)}$$

Define

$$f(x)*f(y) = \frac{f(x)f(y)}{(1-f(x))(1-f(y))+f(x)f(y)} \qquad (3)$$

The assertion of the Lemma follows.

The * operation is a *symmetric summation* [17]. Symmetric summations have the following general structure:

$$\forall x,y \in (0,1): \quad s(x,y) = \frac{g(x,y)}{g(1-x,1-y)+g(x,y)} \qquad (4)$$

where g is the *generating function*. It becomes apparent that in the case of the * operation, g is the product. Since for arguments in the open interval (0,1) the product is positive, continuous, commutative, non decreasing in both arguments and approaches 0 as the arguments approach 0, it follows [17] that the * operation is continuous, commutative and associative in the open interval (0,1).

From equation (1) it is simple to show that
f(-x) = 1 - f(x) and that f(0) = (1/2). This leads to:

$$f(x-x) = f(0) = \tfrac{1}{2}$$
$$f(x-x) = f(x) * f(-x) = f(x) * (1-f(x))$$

i.e. ((0,1), *) is an Abelian group with 1/2 as neutral element and the inverse of an element f(x) is given bei 1- f(x) $\forall$ x $\in$ R.

It should be noticed, that the operation * does not depend on the choice of *k*, but the choice of *k* affects the value of f(x) and f(y) in the former analysis.

It is interesting to mention that the * operation has also been considered by other authors. In [2] it is considered both as an associative symmetric summation and as a symmetric summation generated by the product. An interpretation as a uninorm is given in [4] meanwhile in [12] it is shown that * behaves as a Hamacher-2 t-norm in $[0, \tfrac{1}{2}]^2$ and as a Hamacher-2 dual t-conorm in

$[½, 1]^2$. Finally in [14] * is considered as a non-linear convex combination of the t-norm product and its dual t-conorm.

The above proven Lemma 1 supports the following interpretation. Consider a node in the hidden layer of a neural network with n inputs. The operation of the node may be expressed as:

$$y = f(w_1x_1 + w_2x_2 + ... + w_nx_n)$$

where w_1 , ..., w_n are real-valued input weights to the node. After Lemma 1 the operation at the node may be given by

$$y = f(w_1x_1) * f(w_2x_2) * ... * f(w_nx_n) \qquad (5)$$

Let

$$f_{w_j}(x_j) =_{def} f(w_jx_j) = \frac{1}{1+k^{-w_jx_j}} = \frac{1}{1+(k^{w_j})^{-x_j}}$$

i.e. $f_{w_j}(x_j)$ is a generalized logistic function with parameter k^{w_j}. Then

$$y = f_{w_1}(x_1) * f_{w_2}(x_2) * ... * f_{w_n}(x_n). \qquad (6)$$

From the point of view of fuzzy logic, f_{w_j} may be seen as the characteristic function (fuzzy set) of a fuzzy linguistic term and $f_{w_j}(x_j)$ as the degree of membership of x_j to the fuzzy set f_{w_j}. Since the generalized logistic function represents a "soft step" fuzzy set, the corresponding linguistic term associated to that fuzzy set will be "at least <name of the fuzzy set>" if $w_j > 0$ otherwise "at most <...>". In analogy to the measurements of rise time of the step-answer of amplifiers, the value of x_j leading to $f_{w_j}(x_j) = 0.9$ will be taken as reference.

Accordingly, the fuzzy rule given by a hidden node of the neural network is the following:

if *x_1 is at least T_1* ***and*** *x_2 is at least T_2* ***and...*** ***and*** *x_n is at least T_n* ***then***

$$y = f_{w_1}(x_1) * f_{w_2}(x_2) * ... * f_{w_n}(x_n)$$

where f_{w_j} represents the linguistic term *at least* T_j (assuming that $w_j > 0$).

It becomes apparent, that the output node of the neural network computes then the weighted sum of the former fuzzy rules thus modeling an additive fuzzy system [13].

The effect of a possible bias in a hidden processor upon the interpretation as a fuzzy rule should however still be analyzed. The generalized logistic function turns into:

$$f^{(b)}(x) =_{def} \frac{1}{1+(k)^{-(x-b)}} = f(x - b)$$

From Lemma 1 and from the associativity of * follows that:

$$f(x + y - b) = f(x)*f(y)*f(-b)$$

$$= \frac{f(x)f(y)f(-b)}{(1-f(x))(1-f(y))(1-f(-b)) + f(x)f(y)f(-b)}$$

but since "b" is a constant, $f(-b) = \frac{1}{1+k^b}$ and

$1 - f(-b) = f(b) = \frac{1}{1+k^{-b}} = \frac{k^b}{1+k^b}$. Then:

$$f(x+y-b) = \frac{f(x)f(y)}{k^b(1-f(x))(1-f(y)) + f(x)f(y)} \qquad (7)$$

This equation may well be convenient for computation purposes, but it provides no simple interpretation. The following Lemma however solves this question.

Lemma 2:

$$\forall v \in (0,1) \text{ let} \qquad \lambda^{(b)}(v) =_{def} \frac{v}{v - k^b(v-1)}$$

Then holds:

$$f(x) * f(y) * f(-b) = \lambda^{(b)}(f(x) * f(y)) \qquad (8)$$

It is important to notice that $\lambda^{(b)}$ may be interpreted as a linguistic modifier, acting upon the * operation on the *non-biased* functions f(x) and f(y).

Example:

As an example, the well known test problem "Iris" [3] will be used. This example was processed in [1] to illustrate the translation of the neural network into a set of rules. This will not be repeated here. Only a comparison between a *classical* neural network with three hidden elementary processors with the logistic function as activation function including bias option,

and a neural network with the same architecture but using the generalized logistic function excluding the bias option will be done. In both networks the output node uses a linear activation function. The "Iris" problem is known to have one linear separable class and two overlapping classes, thus being a non-easy problem for feedforward neural networks. By using an extended gradient descent algorithm both weights and k-values were adjusted to minimize the classification error. The obtained values were $k_1 = 1.50544$, $k_2 = 2.44804$ and $k_3 = 1.53952$ and by careful choice of the classification thresholds, (0.35 and 0.78, respectively) *a single* classification error was achieved considering both the training and the test set.

4. CONCLUSIONS

A generalization of the logistic function was introduced and its properties were analyzed. The function computed by each node of the hidden layer may be interpreted as a fuzzy if-then rule such that the resulting premises are aggregated by a symmetric sum. The slope of the generalized logistic function may be adjusted with a gradient descent procedure and this results in modifying the edges of the corresponding fuzzy linguistic terms.

Acknowledgements

The authors are very thankful to Enric Trillas and Ana Pradera for pointing out that the * operator is a symmetric summation and for providing relevant references on the subject. The authors appreciate the suggestions given by one of the Referees to improve the notation used in the paper.

References

[1] Benítez J.M., Castro J.L., Requena I. (1997). Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks* **8**, 1156-1163.

[2] Dubois D., Prade H.(1985). A review of fuzzy set aggregation connectives. *Information Sciences* **36**, 85-121.

[3] Fisher R.A. (1936). The use of multiple measurements in taxonomic problems. *Annual Eugenics* **7**, II:179-188.

[4] Fodor J., Yager R., Rybalov A. (1997). Structure of Uninorms. *Int'l Jr. of Uncertainty, Fuzziness and Knowledge-based Systems*, 411-427.

[5] Han J., Moraga C. (1995). The influence of the sigmoid function parameters on the speed of backpropagation learning. In *From Natural to artificial Neural Computation* (J. Mira, Ed.), *LNCS* **930**, 195-201.

[6] Han J., Moraga C. (1996). Parametric Feedforward Network based Adaptive Fuzzy Modeling. *Proc. Int. Symp. Intelligent Industrial Automation and Soft Computing*, Reading UK, ICSC Academic Press, B–159–165.

[7] Han J., Moraga C., Sinne S. (1996). Optimization of feedforward neural networks. *Engineering Applications of Artificial Intelligence.* **9** (2), 109–119.

[8] Horikawa S.I., Furuhashi T., Uchikawa Y. (1992). A new type of Fuzzy Neural Network for Linguistic Fuzzy Modeling. *Proc. 2nd. Int. Conference on Fuzzy Logic and Neural Networks.* Iizuka, Japan, 1053–1056.

[9] Hornik K., Stinchcombe M., White H. (1989). Multilayer Feedforward Networks are universal approximators. *Neural Networks* **2**, 359–366.

[10] Jang J.S.R. (1993). ANFIS: Adaptive Network based Fuzzy Inference System. *IEEE Trans. on Systems, Man and Cybernetics* **23**, (3), 665–685.

[11] Keller J.M., Tager R.R., Tahani H. (1992). Neural Network implementation of fuzzy logic. *Fuzzy Sets and Systems* **45**, (1), 1–12

[12] Klement P., Mesiar R., Pap E. (1996). On the relationship of associative compensative operators to triangular norms and conorms. *Int'l Jr. of Uncertainty, Fuzziness and Knowledge-based Systems* **4**, (2), 129-144.

[13] Kosko B. (1994). Fuzzy systems as universal approximators. *IEEE Trans. Computers* **43**, (11), 1324–1333,

[14] Mesiar R., Komorniková M. (1998). Triangular norm-based aggregation of evidence under fuzziness. In: *Aggregation and Fusion of Imperfect Information* (B. Bouchon Ed.) *Studies in Fuzziness and Soft Computing* **12**, Physica Verlag, Springer.

[15] Moraga C. (1997). Properties of Parametric Feedforward Networks. *Proceedings XXIII Conferencia Latinoamericana de Informática*, Valparaíso, Chile, 861–870.

[16] Poggio T., Girosi F. (1990). Networks for approximation and learning. *Proceedings of the IEEE* **78**, 1481–1497.

[17] Silvert W. (1979). Symmetric Summation: A class of operations on fuzzy sets. *IEEE Trans. SMC* **9**, 659-667.

[18] Takagi H., Hagashi I. (1991). NN-driven fuzzy reasoning. *Int. Journal of Approximate Reasoning* **5**; (3), 191–212.

[19] Takagi H., Susuki N., Koda T., Kojima Y. (1992). Neural Networks designed on approximate reasoning architecture and their applications. *IEEE Trans. Neural Networks* **3**, (5), 752–760.

[20] Yi H.J., Oh K.W. (1992). Neural Network based Fuzzy Production Rule Generation and its application to an Approximate Reasoning Approach. *Proceedings 2nd. Int. Conference on Fuzzy Logic and Neural Networks*, Iizuka, Japan, 333–336.

A Link-Pruning Algorithm for Neural Networks

J.M. Benítez, J.L. Castro and C.J. Mantas
Dept. Computer Science & Artificial Intelligence (DECSAI)
E.T.S. Ingeniería Informática
Avda. Andalucía, 38
18071 Granada – Spain
e-mail: {J.M.Benitez,castro,cmantas}@decsai.ugr.es

Abstract

This paper proposes a pruning algorithm for Multi-Layered Perceptrons. It is based on an analytical study of the effect of weights on the output. As an straightforward consequence of the Equality Theorem [1], which establishes a direct connection between Artificial Neural Networks and Fuzzy Rule-Based Systems, the pruning algorithm becomes an algorithm to simplify fuzzy rules built with the interactive-or operator.

Keywords: neural network link-pruning, fuzzy rule simplification, approximation error bound

1 Introduction

An usual rule of thumb for achieving good generalization in systems trained with examples is that the smallest system fitting the data should be used. Artificial Neural Networks (ANNs) are systems included in this class, so the rule states that the smallest ANN that models a given unknown system should be employed. The problem is that it is not easy to know what is the *right* size. One common approach to get a model with a size close to the right one is to begin with an ANN large enough to learn the system and then delete parts of it (links and units) which are not necessary. This is done by the so-called pruning algorithms, a survey of which may be found in [8].

Small networks have several advantages besides better expected generalization. They are more efficient. If implemented in hardware, they are cheaper to build. If interested in understanding them, a simpler system is easier to cope with than a bigger one.

The interpretation of ANNs is an important subject, since one of their most criticized aspect is their property of being "black boxes". This problem has received wide attention in the research community. One of the employed approach is exploiting the relationship between ANNs and Fuzzy Rule-Based Systems (FRBSs) [6, 3, 5]. In a previous work [1], the authors have established a strong connection between ANNs and FRBSs. To be more precise, they have shown an equality relationship between Multi-Layered Perceptrons (MLPs) and certain Fuzzy Additive Systems (FASs). This relationship allows to exchange results between both models.

Hence, a pruning algorithm for MLPs also becomes an algorithm for simplifying rule bases in FASs. In the following section a pruning algorithm based on the sensitivity of weights is proposed. Then the corresponding simplification method for fuzzy rules is presented.

2 A pruning algorithm for MLPs

Multilayered feedforward ANNs are especially interesting since they are the most common [4]. Let us consider an ANN with input, hidden and output layers. Let us suppose that the net has n input neurons $(x_1, \ldots, x_n)$, h hidden neurons $(z_1, \ldots, z_h)$, and m output neurons $(y_1, \ldots, y_m)$. Let θ_j be the bias for neuron z_j. Let w_{ij} be the weight of the connection from input neuron x_i to hidden neuron z_j and, β_{jk} the weight of the connection from neuron z_j to output neuron y_k. The function the net calculates is:

$$\mathrm{F} : \mathbb{R}^n \to \mathbb{R}^m; \quad \mathrm{F}(x_1, \ldots, x_n) = (y_1, \ldots, y_m),$$

$$y_k = g_A(\sum_{j=1}^{h} z_j \beta_{jk}), \quad (1)$$

with

$$z_j = f_A(\sum_{i=1}^{n} x_i w_{ij} + \theta_j), \quad (2)$$

where g_A and f_A are activation functions, which are usually continuous, bounded, non-decreasing, non-linear functions. The usual choice is the logistic function: $f_A(x) = \frac{1}{1+e^{-x}}$. Linear functions, even the identity, are also common for g_A.

The pruning algorithm we are to present is based on weight sensitivity calculation. A bound for the error committed in z_j when one of its input links is replaced by a constant value will be used. Such a bound can be obtained as follows.

Let us consider the logistic function applied in the interval $[a + c, b + c]$. Since the logistic function is both continuous and differentiable, the Mean Value Theorem can be applied:

$$f(b + c) - f(a + c) = f'(d)(b - a), \quad (3)$$

with $d \in [a + c, b + c]$.

To get a bound for the above expression, it suffices with bounding the derivative $f'(d)$. But this is immediate, since:

$$f'(x) = f(x)(1 - f(x)), \quad (4)$$

As the maximum for $f'(x)$ is $1/4$, we can write:

$$f(b + c) - f(a + c) \leq \frac{1}{4}(b - a). \quad (5)$$

Now, let us consider the input x_i of the network, which takes values in $[a_i, b_i]$. The logistic function is continuous and differentiable over all $\mathbb{R}$, and hence,

$$f(b_i w_{ij} + \tau) - f(a_i w_{ij} + \tau) \leq \frac{1}{4}|w_{ij}|(b_i - a_i), \quad (6)$$

with:

$$\tau = \sum_{l=1; l \neq i}^{n} x_l w_{lj} + \theta_j.$$

Then, if the term $x_i w_{ij}$ is replaced by

$$\gamma_{ij} = w_{ij} \frac{b_i + a_i}{2}, \quad (7)$$

the following expression holds for all $x_i \in [a_i, b_i]$:

$$|f(x_i w_{ij} + \tau) - f(\gamma_{ij} + \tau)| \leq \frac{1}{8}|w_{ij}|(b_i - a_i). \quad (8)$$

If this bound is smaller than a certain threshold ε, the link from input unit x_i to hidden unit z_j, could be replaced by a constant term, γ_{ij}. The bias term can be adjusted to account for this constant, and the link could be removed rendering a network with an error ε with respect to the complete network.

The effect of this pruning on the k-th output, y_k, is simple to calculate. Firstly, we will consider that the output activation function, g_A is linear. Let z'_j and y'_k be the values obtained for the j-th hidden and k-th output units, respectively, when $x_i w_{ij}$ is replaced by γ_{ij}. Then

$$|y_k - y'_k| = |\beta_{jk}(z_j - z'_j)| \leq \frac{1}{8}|\beta_{jk} w_{ij}|(b_i - a_i). \quad (9)$$

If the sum norm is used to measure the error committed, which is the worst case, the overall error for the network can be bounded as:

$$|y - y'| = \sum_{k=1}^{m} |y_k - y'_k| \leq \frac{1}{8}|w_{ij}(b_i - a_i)| \sum_{k=1}^{m} |\beta_{jk}|. \quad (10)$$

Now, let us consider g_A is the logistic function. Then:

$$\begin{aligned} |y_k - y'_k| &= |f(\beta_{jk} z_j + \eta) - f(\beta_{jk} z'_j + \eta)| \\ &\leq \frac{1}{4}|\beta_{jk}(z_j - z'_j)| \leq \frac{1}{32}|\beta_{jk} w_{ij}(b_i - a_i)|. \end{aligned} \quad (11)$$

and

$$|y - y'| = \sum_{k=1}^{m} |y_k - y'_k| \leq \frac{1}{32}|w_{ij}(b_i - a_i) \sum_{k=1}^{m} |\beta_{jk}|. \quad (12)$$

Expressions (10) and (12) provide a criterion to input-hidden link pruning. If the pruning extends to all the input links to a hidden neuron, the neuron altogether can be deleted, leading to a bigger simplification. Of course, if all output connections from an input node are removed the node itself can be deleted. This way, the pruning process may carry out, as an additional effect, a feature selection task.

The bound is accumulative, so when removing two links the bound for the error committed may be obtained as the sum of the two corresponding individual bounds.

We are now in condition to present the link pruning algorithm. The algorithm takes as inputs a trained neural network and a maximum error bound, ε. It proceeds as follows:

1. For each w_{ij}, calculate $\delta(w_{ij}) := |y - y'|$.
2. Sort the δ in increasing order: $\delta_1, \ldots, \delta_{nh}$.
3. Calculate the first value q such that

$$\sum_{i=1}^{q+1} \delta_i > \varepsilon \qquad \text{and} \qquad \sum_{i=1}^{q} \delta_i \leq \varepsilon. \tag{13}$$

4. The links corresponding to $\delta_1, \ldots, \delta_q$, are the links which will be pruned. Consider every link in turn and readjust the associated bias term.
   - In case all links going to a hidden neuron are deleted, erase the neuron and adjust the bias for the output neurons.
   - In case all links leaving from an input neuron are erased, delete the neuron. This means that input is not relevant to the problem and should not be used.

Once the network has been pruned, its performance can be improved by resuming its training. But this time, its topology will be different, and weights have already been assigned initial values. So the training should be much quicker.

3 Deletion of fuzzy propositions

The study of the relationship between FRBSs and ANNs is an issue on which a number of researchers are spending increasing efforts. Equivalence results linking both models have been shown [6, 3, 5]. In particular, the Theorem of Equality [1, 2] establishes the equality of Three-Layered Perceptrons (as defined by equations (1) and (2)), and a certain class of Fuzzy Additive Systems [7].

The demonstration of the theorem is constructive and transforms the neural network into a FAS with a fuzzy rule per pair of neurons (hidden, output) defined as:

$$R_{jk}: \text{ If } x_1 \text{ is } A^1_{jk} * x_2 \text{ is } A^2_{jk} * \ldots * x_n \text{ is } A^n_{jk} \text{ then } y_k = \beta_{jk}, \tag{14}$$

where $*$ is the interactive-or (i-or, for short) operator defined by [1]:

$$a * b = \frac{a \cdot b}{(1-a)\cdot(1-b) + a \cdot b}. \tag{15}$$

The i-or operator verifies several interesting properties which have a natural interpretation.

Fuzzy sets A^i_{jk} are characterized by their membership function:

$$\mu_{A^i_{jk}}(x) = \frac{1}{1 + \exp(-xw_{ij} - \theta'_j)}, \tag{16}$$

where $\theta'_j = \theta_j / n$. This is the easiest definition for θ'_j, but whatever may be used, as long as the sum of the n θ'_j is θ_j [2].

The transformation on an ANN has an immediate effect on the corresponding FAS. In particular, a simplification of the network implies a simplification of the FRBS.

This way, when a link is removed from the ANN, the corresponding proposition can be removed from the fuzzy rule. Since this proposition is present in one rule per output, it appears exactly m times in the rule base. All of its instances will be erased, it will disappear completely. The function calculated by an ANN and the corresponding FAS is exactly the same, so the bounds calculated in the previous section remain valid when simplifying the fuzzy rule base. All of this can be summarized as: the pruning algorithm for MLP becomes a fuzzy rule base simplifying algorithm.

Whenever a proposition is removed, the constant term replacing it is added to the bias term. The bias is distributed in equals parts among sets A^i_{jk} when building the rules. So, the removal of a proposition and subsequent adjustment implies slight modifications on the remaining fuzzy sets in the proposition.

Within the usual structure of an MLP, we cannot use different values of input-to-hidden weights for different outputs, that is, w_{ij} should have the same value when used to calculate every individual output y_k. In FAS, however, the structure of the fuzzy rules is more flexible, since the contribution of a weight to each output is distributed over different, independent rules. This can

be easily check in the samples included in [1, 2], where every weight (from input to hidden layer) w_{ij} appears in m rules, one per output. Each of these rules can be used individually, so it is possible to delete proposition "x_i is A^i_{jk}" in rule R_{jk}, and keep it in rule $R_{jk'}$.

The previous observation leads to a slightly different simplifying algorithm for the fuzzy rule bases of these FASs. This algorithm employs (9) to evaluate the error committed when replacing a proposition by a constant, instead of (10). More explicitly, the algorithm takes as inputs the fuzzy rule base $\{R_{jk} : 1 \leq j \leq h, 1 \leq k \leq m\}$ and a maximum error bound, ε. It proceeds as follows:

1. For each proposition $p_{ijk} \equiv$"x_i is A^i_{jk}", calculate $\delta(p_{ijk})$:

$$\delta(p_{ijk}) = \frac{1}{8}|\beta_{jk} w_{ij}|(b_i - a_i). \qquad (17)$$

2. Sort the δ in increasing order: $\delta_1, \ldots, \delta_{nhm}$.

3. Calculate the first value q such that

$$\sum_{i=1}^{q+1} \delta_i > \varepsilon \qquad \text{and} \qquad \sum_{i=1}^{q} \delta_i \leq \varepsilon. \qquad (18)$$

4. The propositions corresponding to $\delta_1, \ldots, \delta_q$, are the propositions which will be removed. Consider every proposition in turn and readjust the remaining propositions in the rule.

   - In case all propositions in a rule are removed, the rule is replaced by a constant term.
   - If all propositions regarding an input variable x_i are deleted, the input is not relevant to the problem.

4 Conclusions and final remarks

In this paper, we have studied the effect of weights on outputs of a three-layered perceptron. This allows us to calculate a bound for the error committed when replacing a link by a constant term. A link-pruning algorithm has been proposed, which based on the obtained bound, removes links of minor effect in the network. The deletion of several links may induce the elimination of hidden, or even, input units.

By applying the Theorem of Equality between MLPs and a wide class of FASs, the previous link-pruning algorithm becomes a fuzzy rule simplification algorithm for these FRBSs. Due to the more flexible organization of a rule base, the simplification algorithm can be modified to remove individual propositions, independently from a rule to another.

Acknowledgements

This research has been partially supported by CICYT under project TIC96-1393-C06-04.

References

[1] J.M. Benítez, J.L. Castro, and I. Requena. Are artificial neural networks black boxes? *IEEE Trans. Neural Networks*, 8(5):1156–1164, September 1997.

[2] J.M. Benítez. *Extraction of Fuzzy Rules with and from Artificial Neural Networks.* PhD thesis, Dept. Computer Science & Artificial Intelligence, University of Granada, Granada, Spain, 1998. (In Spanish).

[3] J.J. Buckley, Y. Hayashi, and E. Czogala. On the equivalence of neural nets and fuzzy expert systems. *Fuzzy Sets and Systems*, 53:129–134, 1993.

[4] S. Haykin. *Neural Networks. A comprehensive foundation.* Prentice-Hall, 1999.

[5] K.J. Hunt, R. Haas, and R. Murray-Smith. Extending the functional equivalence of radial basis function networks and fuzzy inference systems. *IEEE Trans. Neural Networks*, 7(3):776–781, May 1996.

[6] J.-S. R. Jang and C.-T. Sun. Functional equivalence between radial basis function networks and fuzzy inference systems. *IEEE Trans. Neural Networks*, 4(1):156–158, 1992.

[7] B. Kosko. Fuzzy systems as universal approximators. *IEEE Trans. Computers*, 43(11):1324–1333, 1994.

[8] R. Reed. Pruning algorithms—A survey. *IEEE Trans. Neural Networks*, 4(5):740–747, September 1993.

FUZZY PATH TRACKING OF AUTONOMOUS VEHICLES USING GPS

A. Rodriguez-Castaño, G. Heredia and A. Ollero.

Departamento de Ingeniería de Sistemas y Automática.
Escuela Superior de Ingenieros, Universidad de Sevilla.
Camino de los Descubrimientos, 41092 Sevilla (Spain).
Fax:+34-95-4487340. Email: guiller@cartuja.us.es

Abstract

This paper presents an autonomous vehicle position estimation system based on GPS, that uses a fuzzy sensor fusion technique. A fuzzy path tracking algorithm is also proposed. Both systems have been implemented in the ROMEO-4R vehicle developed at the University of Seville.

Keywords: Fuzzy position estimation, GPS, fuzzy path tracking, autonomous vehicles

1 INTRODUCCION

Path tracking of autonomous vehicles consists in the generation of the steering commands for a vehicle to follow autonomously a given path taking into account the constraints imposed by the vehicle and the lower level motion controllers. Many path tracking strategies have been presented. In the experiments described in this paper a fuzzy control technique is applied.

Position estimation is the process of calculating the position and orientation of the mobile robot with respect to a given reference system. Position estimation is very important for path tracking. If the estimation is not sufficiently accurate, oscillations and erratic behavior of the vehicle may occur.

The Global Positioning System (GPS) provides a good source of absolute position information for autonomous navigation, and it can be used for primary position estimation in many cases. However, the receiver can give erroneous measures in certain situations, and therefore, internal sensors as encoders, gyroscopes and compasses have to be used in conjunction with the GPS [4]. The readings of all these sensors are combined using sensor fusion techniques [6]. In this paper, a fuzzy sensor fusion technique has been used.

This paper first presents the path tracking problem. Section 3 introduces the Global Positioning System, and its use for mobile robot position estimation. Section 4 presents the fuzzy sensor fusion adopted to deal with the GPS signal degradations. In Section 5, a direct fuzzy controller for the path tracking problem is presented. In Section 6, field experiments with the ROMEO-4R autonomous vehicle are presented. The last three Sections are for the conclusions, acknowledgments and references.

2 PATH TRACKING

Several path tracking algorithms select a goal point P along the path at a fixed distance from the vehicle that is called the lookahead distance L (see Figure 1). Using this goal point three values are obtained that become the inputs to the path tracker. These inputs are the lateral position error ξ in vehicle coordinates (x,y), the orientation error with respect to the goal point θ_e and the curvature error with respect to the goal point γ_e.

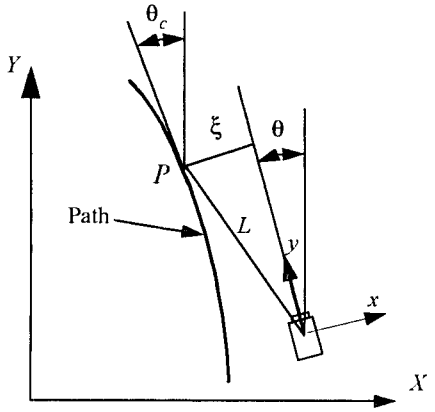


Figure 1 Path tracking

The expressions of θ_e and γ_e are:

$$\theta_e = \theta - \theta_c \qquad \gamma_e = \gamma - \gamma_c \tag{1}$$

where θ_c and γ_c are the orientation and curvature of the path at the goal point.

Direct fuzzy control [1] uses a fuzzy controller that has as inputs the lateral displacement of the vehicle from the goal point on the desired path, the deviation angle from the goal point and the deviation in curvature. The output of the fuzzy controller is the steering command given to the vehicle.

3 POSITION ESTIMATION USING GPS

The Global Positioning System (GPS) is a satellite navigation system capable of providing a highly accurate, continuous global navigation service independent of other positioning aids.

The system uses de NAVSTAR satellites which consist of 21 operational satellites and 3 active spares to provide a GPS receiver with a 6 to 10 satellite coverage at all times.

GPS civilian code is capable of absolute position accuracies of about 100 metres or less. This level of accuracy is clearly insufficient for mobile robot position estimation, and it is due to many factors: atmospheric delays, ephemeris errors, satellite and receiver clock errors, multipath signal reception and selective availability. There is a technique known as Differential GPS which eliminate or greatly reduce most of these errors.

Differential operation requires that stations operate in pairs: a monitor in a fixed, known position, and a remote station onboard the mobile robot. Both stations must be connected with a radio link. The monitor station computes the errors between the position obtained from the GPS satellites signals and its known position, and transmits these corrections to the remote station through the radio link. Differential GPS receivers are capable of real time accuracies of a few centimeters.

The main problem with GPS is that the receivers need to be in direct sight of satellites, and thus periodic signal blocking occurs because of buildings, foliage and hilly terrain. Another problem is that corrections from the radio link may arrive "corrupted". Most DGPS receivers can provide an estimation of the accuracy level of the calculated position: centimeter, submeter and meter accuracy. This accuracy level can be used to decide whether to rely on internal sensors (encoders and gyroscopes).

Accuracy in GPS is a statistical measure of performance. A few tests with a Trimble 7400 Msi have shown that even with an error less than 2 cm 99% of the time, a few points with a meter or tens of meters error can be found. These points must be filtered, as will be seen in the next Section.

4 FUZZY SENSOR FUSION

In autonomous navigation, the measures of several redundant sensors are normally used for position estimation, due to noisy readings of individual sensors. To obtain an estimate of the position these measures have to be combined, what is called "sensor fusion". Kalman filtering is the most widely used sensor fusion technique in autonomous navigation.

Readings of an individual sensor may be erroneous, for example, the GPS receiver may loose contact with the satellites, or the compass may give an erroneous measure due to a local variation of the earth magnetic field. In these cases, the Kalman filter cannot deal satisfactorily with these measures. This paper proposes the application of a fuzzy sensor fusion technique to consider the heuristic knowledge involved in the estimation problem. This technique is based on the use of a fuzzy system that filters these erroneous points.

The fuzzy system can be split up into two subsystems, a position fuzzy system and an orientation fuzzy system. Then, the position is computed by means of:

$$\text{Pos} = \text{GPS_position} * \text{pos_GPS_factor} + (1 - \text{pos_GPS_factor}) * \text{DR_position}$$

where pos_GPS_factor is the position fuzzy system output, which gives the confidence in the sensors. DR_position is the vehicle position computed using dead reckoning:

$$\begin{aligned} x(k+1) &= x(k) - v(k)\Delta T \sin\theta(k) \\ y(k+1) &= y(k) + v(k)\Delta T \cos\theta(k) \end{aligned} \tag{2}$$

where $v(k)$ and $\theta(k)$ are respectively the vehicle velocity and orientation in the time k, and ΔT is the sampling period.

The position fuzzy system inputs are the level of accuracy (meter, submeter, centimeter), the distance travelled by the DR estimation since last GPS update and the position error between the DR estimation and the GPS position. The rules are as:

IF accuracy IS high AND pos_error IS low THEN pos_GPS_factor IS high
IF accuracy IS high AND pos_error IS high AND dist_travelled IS low THEN pos_GPS_factor IS low
IF accuracy IS low THEN pos_GPS_factor IS low

where the membership functions of the inputs and the output are defined in Figure 2.

A similar scheme is used in the orientation fuzzy system. The orientation is given by:

or = or_compass * compass_factor + or_gyro * gyro_factor + or_GPS * or_GPS_factor

where the factors give the confidence in the sensors, and the following equation is used to compute the orientation from the gyroscope:

$$\theta(k+1) = \theta(k) + \omega(k)\Delta T \tag{3}$$

where $\theta(k)$ and $\omega(k)$ are respectively the orientation and the yaw rate, and ΔT is the sampling period.

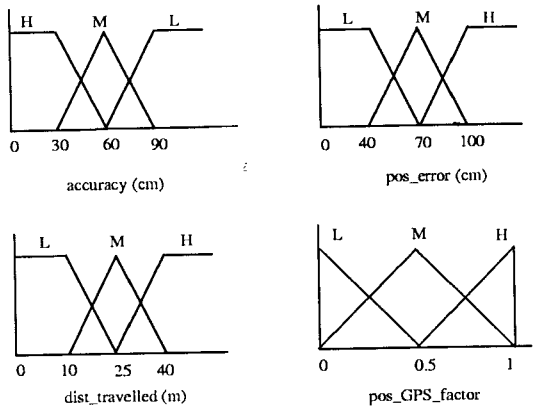


Figure 2 Membership functions

The orientation fuzzy system inputs are the yaw rate, the orientation estimated from the GPS (or_GPS), the orientation computed from the gyroscope (or_gyro) and the orientation given by the compass (or_compass). The ouputs of the system are compass_factor, gyro_factor and or_GPS_factor. The rules are similar to the ones presented before.

5 DIRECT FUZZY CONTROLLER

Consider the direct fuzzy feedback control scheme shown in Figure 3. The objective is that the vehicle tracks a given explicit path as shown in Figure 1. The path is defined by the position, orientation and curvature of their points. An objective point is selected on the path at a given lookahead distance L, which is a parameter of the controller.

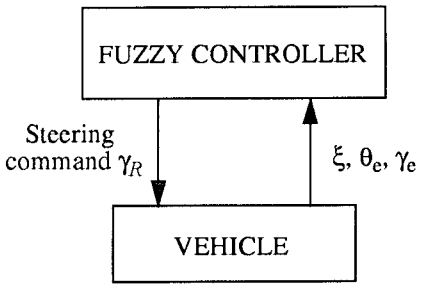


Figure 3 Fuzzy path tracking

The fuzzy controller has as inputs the lateral deviation ξ from the goal point (local x coordinate in a system attached to the vehicle), the deviation in heading θ_e, and the deviation in curvature γ_e.

The output of the fuzzy controller is the required steering command γ_R. The heuristic to drive the vehicle from the deviations is very simple. Assuming that positive steering is counterclockwise, the rules are as:

R_i: if ξ is POSITIVE LARGE, θ_e is NEAR ZERO and γ_e is POSITIVE, then γ_{Ri} is POSITIVE SMALL

where the membership functions of the inputs (ξ, θ_e, γ_e) and the output (γ_R) are defined in Figure 4. Five linguistic terms are used for the output, while for the three inputs 7, 5 and 3 linguistic terms have been defined respectively.

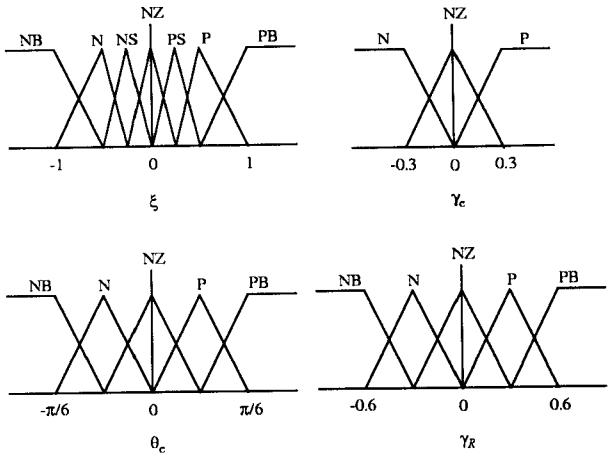


Figure 4 Membership functions

Some other heuristic consideration to take into account preferences to solve conflicts between rules are used (see for example [1]).

The above rules define a Mamdani-type controller. It is also possible to use a Takagi-Sugeno type controller for path tracking [2], with rules like:

R_j: if ξ is POSITIVE, θ_e is NEAR ZERO and γ_e is NEGATIVE, then $\gamma_{Rj} = 0.2\ \xi - 0.1\ \theta_e$

In [2] a method to compute TS fuzzy controllers using control theory and fuzzy identification procedures is presented.

It is also possible to identify the TS fuzzy controller by learning from human drivers. Particularly, fuzzy identification has been applied to compute the controller from the recording of input/output data (ξ, θ_e, γ_e ; γ_R) when the vehicle is steered by a human driver when tracking a path. Two different recording techniques have been implemented: a) an operator guides the robot by means of a joystick [1], and b), the vehicle admits a human driver on board, which drives the vehicle (with the automatic system switched off) [5].

The use of fuzzy logic methods for the implementation of path tracking controllers provides the following benefits:

- The heuristic driving knowledge can be incorporated using few simple and understandable rules.

- Fuzzy control techniques could interpolate the rules to obtain smooth control surfaces.
- Learning procedures to compute the fuzzy controller from experiments when a human is driving the vehicle can also be applied.
- Local control laws computed by means of control theory can be embedded in the fuzzy control strategy (see, for example, [3]).

6 FIELD EXPERIMENTS

Several field experiments have been done with the ROMEO-4R autonomous vehicle, developed at the Departamento de Ingeniería de Sistemas y Automática of the University of Seville [3].

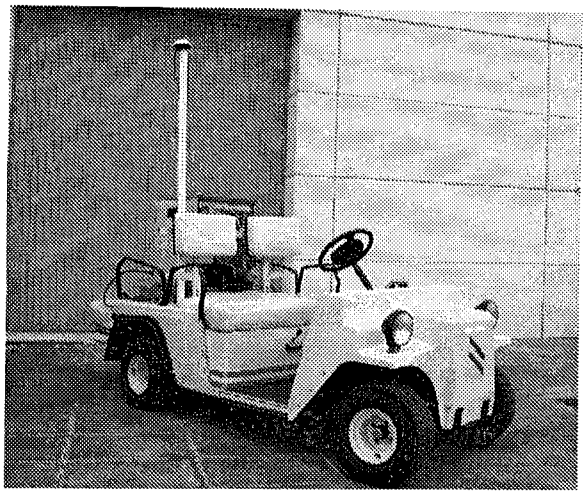


Figure 5 ROMEO-4R autonomous vehicle

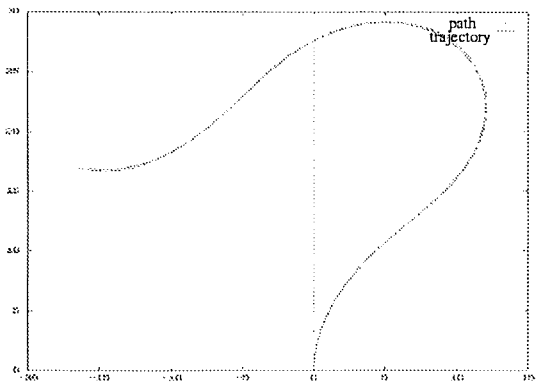


Figure 6 Trajectory followed by ROMEO-4R in a path tracking experiment

ROMEO-4R is a four-wheeled vehicle with Ackerman steering (see Figure 5). Automatic steering is implemented by using a DC motor that is connected to the steering wheel axle through a reduction gear. ROMEO-4R carries a Differential GPS system for primary position estimation. The receiver, a RT-20 model from Novatel, is capable of achieving accuracies of 20 cm or less. The vehicle also has optical encoders for motion control and dead-reckoning, gyroscopes and a compass for the orientation.

In Figure 6, the trajectory followeed by the ROMEO-4R in a path tracking experiment with the direct fuzzy path tracker described in this paper, and using DGPS as the primary positioning system.

7 CONCLUSIONS

Fuzzy control techniques have proved efficiency for path tracking techniques of autonomous vehicles. In this paper the results of the application of these techniques when using GPS is presented. The paper includes a new fuzzy logic position estimation method which integrates the data from the GPS with the measurement obtained from a compass, a gyro and the encoders in the vehicle wheels. The method is very reliable to failures in the sensors including temporary loss of the GPS signal.

Acknowledgments

This work has been done in the CICYT Project TAP96-1184-C04-01 and the FAMIMO Project, (European Commission, ESPRIT LTR 21911).

References

[1] García-Cerezo, A., A. Ollero and J. L. Martínez, 1996, "Design of a robust high performance fuzzy path tracker for autonomous vehicles, *Int. Journal of Systems Science*, vol. 27, n. 8, pp. 799-806.

[2] Ollero, A., A. García-Cerezo and J. L. Martínez, 1996, "Design of fuzzy controllers from heuristic knowledge and experimental data", *1996 IFAC World Congress*, paper 1d-01 5, Vol. A, pp. 433-438.

[3] Ollero, A., B.C. Arrúe, J. Ferruz, G. Heredia, F. Cuesta, F. López-Pichaco and C. Nogales, 1998. Intelligent Components in the Romeo Vehicles *Proc. of the IFAC Workshop on Intelligent Componenets for Vehicles (ICV'98)*, Sevilla, Spain, March 23-24.

[4] Rintanen, K., H. Makela, K. Koskinen, J. Puputti, M. Sampo, M. Ojala, 1995, Development of an autonomous navigation system for an outdoor vehicle and inertial navigation, *Proc. of the 2nd. Int. Conf. on Intelligent Autonomous Vehicles (IAV95)*, Helsinki, Finland, pp. 220-225.

[5] Sánchez, O., A. Ollero and G. Heredia, 1997. Adaptive Fuzzy Control for Automatic Path Tracking of Outdoor Mobile Robots. Application to Romeo 3R. *Proc. of the 6th IEEE Int. Conf. on Fuzzy Systems (FUZZ-IEEE'97)*, Barcelona, Spain, July 1-5.

[6] Schonberg, T., M. Ojala, J. Suomela, A. Torpo and A. Halme, 1995, Positioning an autonomous off-road vehicle by using fused DGPS and inertial navigation, *Proc. of the 2nd. Int. Conf. on Intelligent Autonomous Vehicles (IAV95)*, Helsinki, Finland, pp. 226-231.

THE APPLICATION OF GENERALISED CONSTRAINTS TO OBJECT-ORIENTED DATABASE MODELS

Guy de Tré
University of Ghent
Computer Science Laboratory
Sint-Pietersnieuwstraat 41,
B-9000 Ghent, Belgium
guy.detre@rug.ac.be

Rita de Caluwe
University of Ghent
Computer Science Laboratory
Sint-Pietersnieuwstraat 41,
B-9000 Ghent, Belgium
rita.decaluwe@rug.ac.be

Summary

This paper presents a formal framework for a generalised object-oriented database model that is able to cope with fuzzy and uncertain information. The generalised model is obtained as a generalisation of a crisp object-oriented database model that is compliant with the ODMG de facto standard and is built upon an algebraic type system and a constraint system.

Keywords: Object-oriented database models, generalised constraints, fuzzy and uncertain information management.

1 Introduction

During the past decade several "fuzzy" object-oriented database models have been proposed. An overview can be found in [2]. These models do not conform to a single underlying object data model, due to a lack of object standards.

The ODMG de facto standard data model [1] offers new promising perspectives. However, it still suffers from several shortcomings such as some lack of formal semantics and its limited ability to deal with constraints, despite the fact that a thorough support of constraints is the most obvious way to guarantee the integrity of a database. The formal framework we present is ODMG compliant and overcomes the mentioned ODMG shortcomings. It also supports both fuzziness and uncertainty.

By building a formal generalised type system and constraint system we can give the formal definition of a generalised object and a generalised database scheme, both of them being the main components of our generalised object-oriented database model.

2 A formal framework for generalised object-databases

Since crisp databases can be seen as special cases of fuzzy and/or uncertain databases, it is our aim to define a generalised object-oriented database model which is a generalisation of a crisp one (rather than defining an extension of a crisp database model).

2.1 The definition of types

Types are defined as instances of a generalised type system. Our definition of a generalised type system is the generalisation of the definition of a crisp type system which is in accordance with the directions given in [4] and completely compliant with the ODMG specifications.

2.1.1 The crisp type system

In our type system, we state the rules which define literal types, object types, reference types and the **Void** type. A literal type as either a base type, a collection literal type or a structured literal type. The set of the type specifications of all the literal types will be denoted as $T_{literal}$, the set of the specifications of all the object types as T_{object}, and the set of the specifications of all the reference types as $T_{reference}$.

We do not limit ourselves to the syntax of the type specifications, but also endow the definition rules with formal semantics. Our semantic definitions are related to domains and sets of domains. The semantics of a type specification ts are completely captured by a set of domains D_{ts}, a designated domain $dom_{ts} \in D_{ts}$, a set of operators O_{ts} and a set of axioms A_{ts}. Each domain contains an "undefined" value $\perp_{ts}$.

The (crisp) type system TS, which allows to derive both the specification and the implementations of the valid types, is defined as the triple:

$$TS = [T, P, f_{impl}^{type}]$$

where $T = \{\textbf{Void}\} \cup T_{reference} \cup T_{literal} \cup T_{object}$ is the set of valid type specifications, P is the infinite set of all programs —indepedent of the language in which they are written— and f_{impl}^{type} is called the (type) implementation function:

$$\begin{aligned} f_{impl}^{type} : T &\rightarrow \wp(P) \\ ts &\mapsto \{p_1, \ldots, p_n\} \end{aligned}$$

f_{impl}^{type} is a mapping which maps each type specification ts of the domain T onto the subset $\{p_1, \ldots, p_n\}$ of its co-domain $\wp(P)$ (the powerset of P) that contains all the implementations of ts.

An instance t of the type system TS is called a type and is defined by a couple:

$$t = [ts, f_{impl}^{type}(ts)], \text{ with } ts \in T$$

When $ts = \textbf{Void}$, t is called a void type; when $ts \in T_{literal}$, t is a literal type; when $ts \in T_{object}$, t is an object type; otherwise t is called a reference type.

Here, we focus on the object types, as they provide a formal basis for the most important notion in ODMG, namely the object. The syntax of the specification $ts \in T_{object}$ of an arbitrary object type t is defined as:

$$\textbf{Class } id : \widehat{id}_1, \ldots, \widehat{id}_m(id_1 : s_1; \ldots; id_n : s_n)$$

where id is an identifier that represents the name of the object type, $\widehat{id}_i, 1 \leq i \leq m$ are identifiers representing the parent types of t (if any) and $id_i : s_i, 1 \leq i \leq n$ are the characteristics (attributes, relationships and methods) which are explicitly specified within the syntax of the object type. For each $id_i : s_i, 1 \leq i \leq n$, id_i is the identifier and s_i is the specification of the characteristic, which for attributes and relationships is a type specification and for methods is a signature.

Example 1: With **String**, **Integer**, **Set** and the enumeration type **Enum** $TLang(Dutch, French, English)$ being an element of $T_{literal}$, the specification ts of an object type $TPerson = [ts, f_{impl}^{type}(ts)]$ can be given as:

$$\begin{aligned} \textbf{Class } TPerson(\quad & Name : \textbf{String}; \\ & Age : \textbf{Integer}; \\ & Languages : \textbf{Set}(TLang)) \end{aligned} \quad \diamond$$

The properties as well as the behaviour of both objects and literals are formalised by the types of our type system. (As far as the literals are concerned, the definition of the behaviour is obviously limited to the behaviour which is implicitly defined.) An object o or instance of an object type is defined as a quadruple

$$o = [oid, N, t, v]$$

where oid is a unique object identifier, N is a set of object names, $t = [ts, f_{impl}^{type}(ts)]$ is the object type, and $v \in dom_{ts}$ is the state of the object. N can be an empty set and for every attribute or relationship $id_i : s_i$ specified within t, v contains a value $v_i \in dom_{s_i}$. The extent of an object type is the set of all its (persistent) instances. If an object is an instance of type t, then it has to be a member of the extent of t. If type t is a subtype of type $\widehat{t}$, then the extent of t is a subset of the extent of $\widehat{t}$.

Example 2: An instance of $TPerson$ is, e.g.

$$\begin{aligned} [oid1, \emptyset, \quad & TPerson, \textbf{Struct}(\text{“Ann”}, \\ & 28, \textbf{Set}(Dutch, French, English))] \end{aligned} \quad \diamond$$

2.1.2 The generalised type system

A generalised (fuzzy) type system is obtained by generalising the definition of the crisp type system

$$TS = [T, P, f_{impl}^{type}]$$

First of all, the set of type specifications T is generalised to a set $\tilde{T}$ by generalising the definitions of the domains and operators of the type specifications $ts \in T$. For every type specification $ts \in T$, the generalised counterpart $\tilde{ts}$ has a domain $dom_{\tilde{ts}}$ that is defined as the set

$$\tilde{\wp}(dom_{ts})$$

of fuzzy sets on dom_{ts} and a set of operators $O_{\tilde{ts}}$ that contains the generalised counterparts of the operators of O_{ts} (generalised using Zadeh's extension principle).

Furthermore, the definition of a characteristic $id_i : s_i$ of an object type is generalised through the use of generalised constraints as defined by L.A. Zadeh [6]: a generalised attribute or relationship is defined as

$$\tilde{id}_i \; isr \; \tilde{ts}_i$$

where isr is a variable copula and r is a discrete variable which value defines the way in which the values of the (domain of the) characteristic are constrained. For now, the considered types of constraints are:

1. *Equality constraint*, $r = e$. In this case $\tilde{id}$ *ise* $\tilde{ts}$ means that $\tilde{id}$ will be assigned a crisp value of $dom_{\tilde{ts}}$. This case is semantically equivalent with the definition $id : ts$ of its crisp counterpart.
2. *Possibilistic constraint*, $r = blank$. In this case $\tilde{id}$ *is* $\tilde{ts}$ means that $\tilde{id}$ will be assigned a value of $dom_{\tilde{ts}}$, i.e. a fuzzy set, and $\tilde{id}$ is seen as a disjunctive (possibilistic) variable.
3. *Veristic constraint*, $r = v$. In this case, $\tilde{id}$ *isv* $\tilde{ts}$ means that $\tilde{id}$ will be assigned a value of $dom_{\tilde{ts}}$, i.e. a fuzzy set, and $\tilde{id}$ is seen as a conjunctive (veristic) variable.

With these generalisations, the generalised type system GTS is defined as the triple:

$$GTS = [\tilde{T}, P, \tilde{f}^{type}_{impl}]$$

where $\tilde{T}$ is the generalised set of valid type specifications, P remains the infinite set of all programs and $\tilde{f}^{type}_{impl}$ is a mapping which maps each generalised type specification $\tilde{ts}$ of $\tilde{T}$ onto its set of implementations $\tilde{ts}$:

$$\begin{aligned} \tilde{f}^{type}_{impl} : \tilde{T} &\rightarrow \wp(P) \\ \tilde{ts} &\mapsto \{p_1, \ldots, p_n\} \end{aligned}$$

An instance $\tilde{t}$ of the generalised type system GTS is called a generalised type and is defined by a couple:

$$\tilde{t} = [\tilde{ts}, \tilde{f}^{type}_{impl}(\tilde{ts})], \text{ with } \tilde{ts} \in \tilde{T}$$

Example 3: With **String**, **Integer** and the enumeration type **Enum** $TLang(Dutch, French, English)$ being an element of $\tilde{T}$, the specification $\tilde{ts}$ of a (generalised) object type $GTPerson = [\tilde{ts}, \tilde{f}^{type}_{impl}(\tilde{ts})]$ can be given as:

Class $GTPerson($ $Name\ ise$ **String**;
$Age\ is$ **Integer**;
$Languages\ isv\ TLang)$ $\diamond$

The generalisation of the definition of an object becomes:

$$\tilde{o} = [oid, N, \tilde{t}, \tilde{v}, Pos, Nec]$$

where oid remains a unique object identifier, N remains a set of object names, $\tilde{t} = [\tilde{ts}, \tilde{f}^{type}_{impl}(\tilde{ts})]$ is the generalised object type, and the state $\tilde{v}$ is now an element of $dom_{\tilde{ts}}$. Additionally, a generalised object has a (system defined) possibility measure Pos and a necessity measure Nec, which can take values between 0 and 1, and which together express the fuzzy thruth value

$$\{Pos/true, Nec/False\}$$

of the object as an instance of the generalised object type $\tilde{t}$.

Example 4: An instance of the generalised object type $GTPerson$ is, e.g.

$[oid1$ $, \emptyset, GTPerson,$ **Struct**("Ann",
$\{.6/26, .8/27, 1/28, 1/29, .8/30, .6/31\},$
$\{.4/Dutch, .7/French, 1/English)), 1, 0]$ $\diamond$

2.2 The definition of constraints

Constraints are used to enforce integrity rules on databases (e.g. domain rules, referential integrity, etc.) and to help to specify the formal semantics of a database model (e.g. null values, definitions of keys, etc.) [3]. In our approach a constraint is formally defined as an instance of a constraint system.

2.2.1 The crisp constraint system

A crisp constraint is defined as a Boolean function, which indicates whether a given object is valid within the context of a given database or not. The (crisp) constraint system CS, which allows to derive both the specification and the implementations of the constraints, is defined as the triple:

$$CS = [C, P, f^{constr}_{impl}]$$

where C is the set of the specifications of all the valid constraints, and P and f^{constr}_{impl} denote respectively the infinite set of all programs and the (constraint) implementation function, which is defined as:

$$\begin{aligned} f^{constr}_{impl} : C &\rightarrow \wp(P) \\ cs &\mapsto \{p_1, \ldots, p_n\} \end{aligned}$$

Hereby, f^{constr}_{impl} is a mapping which maps each constraint specification cs of the domain C onto the subset $\{p_1, \ldots, p_n\}$ of the co-domain $\wp(P)$ (the powerset of P) that contains all the implementations of cs.

An instance c of the constraint system CS is called a constraint and is defined by a couple:

$$c = [cs, f^{constr}_{impl}(cs)], \text{ with } cs \in C$$

2.2.2 The generalised constraint system

The generalisation of the constraint system is straightforward: a Boolean (constraint) function can be generalised to a function with co-domain $[0, 1]$ that associates a membership degree with a given object, indicating to which degree the object satisfies the constraint. As well the constraints which specify the domain rules, as the constraints which specify the transition rules are generalised in this way, which leads to the definition of the generalised constraint system GCS.

$$GCS = [\tilde{C}, P, \tilde{f}^{constr}_{impl}]$$

where $\tilde{C}$ is the generalised set of valid constraints specifications, P is the set of all programs and $\tilde{f}^{constr}_{impl}$ denotes the generalised (constraint) implementation function:

$$\begin{aligned} \tilde{f}^{constr}_{impl} : \tilde{C} &\rightarrow \wp(P) \\ \tilde{cs} &\mapsto \{p_1, \ldots, p_n\} \end{aligned}$$

2.3 The definition of generalised object schemes and database schemes

The adopted scheme definitions are the counterparts of the ones presented in [5] for the relational database model. Both the definitions of generalised object scheme and generalised database scheme rely on the definitions of types and constraints.

2.3.1 A generalised object scheme

The full semantics of an object are described by a generalised object scheme $\tilde{os}$. This scheme defines all the characteristics of the object, including the constraints that apply to the object and is defined by an object type $\tilde{t}$ (i.e. an instance of the generalised type system GTS), a meaning $\tilde{M}$ and a conjunctive fuzzy set of constraints $\tilde{C}_{\tilde{t}}$ (i.e. a finite fuzzy set over the instances of the generalised constraint system GCS).

$$\tilde{os} = [\tilde{t}, \tilde{M}, \tilde{C}_{\tilde{t}}]$$

The meaning $\tilde{M}$ is an informal component of the definition since it will mostly be described in a natural language [5]. The membership degree of an element of $\tilde{C}_{\tilde{t}}$ indicates to which degree the constraint applies to the object type $\tilde{t}$.

An instance of the object type $\tilde{t}$ is defined to be an instance of the generalised object scheme $\tilde{os} = [\tilde{t}, \tilde{M}, \tilde{C}_{\tilde{t}}]$ if and only if it satisfies (with a non-zero membership degree) all the constraints of $\tilde{C}_{\tilde{t}}$ and all the constraints of the sets $\tilde{C}_{\tilde{\tilde{t}}}$ of the object schemes $[\tilde{\tilde{t}}, \tilde{\tilde{M}}, \tilde{C}_{\tilde{\tilde{t}}}]$ which have been defined for the supertypes $\tilde{\tilde{t}}$ of $\tilde{t}$.

The possibility measure Pos and the necessity measure Nec result from the degrees to which the object satisfies the constraints of $\tilde{C}_{\tilde{t}}$ (by calculating t-conorms and t-norms).

2.3.2 A generalised database scheme

A generalised database scheme $\tilde{ds}$ describes the information which is stored in the generalised database and is defined as a triple:

$$\tilde{ds} = [\tilde{D}, \tilde{M}, \tilde{C}_{\tilde{D}}]$$

where $\tilde{D} = \{\tilde{os}_i = [\tilde{t}_i, \tilde{M}_i, \tilde{C}_{\tilde{t}_i}] | 1 \leq i \leq n, i, n \in N_0\}$ is a finite set of generalised object schemes, $\tilde{M}$ represents the meaning of $\tilde{ds}$, and $\tilde{C}_{\tilde{D}}$ is a conjunctive fuzzy set of constraints which imposes extra conditions on the instances of the set of object schemes (e.g. referential constraints between two object schemes). Again, the membership degrees are an indication for the relevance of the constraints. Every generalised object scheme in $\tilde{D}$ is defined for a different object type. If a generalised object scheme $\tilde{os}_i \in \tilde{D}$ is defined for an object type $\tilde{t}$ and $\tilde{t}'$ is a supertype of $\tilde{t}$ or $\tilde{t}'$ is related with $\tilde{t}$, then there exists a generalised object scheme $\tilde{os}'_i \in \tilde{D}$ which is defined for $\tilde{t}'$.

An instance of a generalised object scheme $\tilde{os}_i \in \tilde{D}$ is defined to be an instance of the generalised database scheme $\tilde{ds} = [\tilde{D}, \tilde{M}, \tilde{C}_{\tilde{D}}]$ if and only if it satisfies all the constraints of $\tilde{C}_{\tilde{D}}$. All the instances of the generalised database scheme $\tilde{ds}$ are elements of the extent of $\tilde{ds}$.

As is the case for the instances of an object scheme, the measures Pos and Nec result from the degrees to which the object satisfies the constraints of $\tilde{C}_{\tilde{D}}$ (by calculating t-conorms and t-norms).

2.4 The definition of a (fuzzy) database

The extent of a generalised database scheme $\tilde{ds}$ is called a (fuzzy) database and is by definition a set of conjunctive fuzzy sets of database instances (grouped per object scheme). For each database instance

$$\tilde{o} = [oid, V_{names}, \tilde{t}, \tilde{v}, Pos, Nec]$$

the membership degree of the instance within its conjunctive fuzzy set is calculated as $(Pos + 1 - Nec)/2$ and can be seen as the fuzzy thruth value of the object in the database.

3 Conclusions

A formal framework for the definition of a fuzzy and/or uncertain object-oriented database model has been presented. This framework is based on a type system and a related constraint system, which is meant to guarantee database integrity. Zadeh's extension principle and generalised constraints have been used to generalise the database model. Subsequently, databases have been defined as sets of fuzzy sets of objects.

References

[1] Cattell, R. G. G. et al. (1997). *The Object Database Standard: ODMG 2.0.* Morgan Kaufmann Publishers Inc., San Francisco, CA USA.

[2] De Caluwe, R. (ed.) (1997). *Fuzzy and Uncertain Object-Oriented Databases: Concepts and Models.* Advances in Fuzzy Systems - Applications and Theory, Vol. 13, World Scientific, Singapore.

[3] De Tré, G. and De Caluwe, R. (1999). *A generalised object-oriented database model with generalised constraints.* Proceedings of the NAFIPS'99 conference, IEEE, New York, NY, pp. 381–386.

[4] Lausen, G. and Vossen, G. (1998). *Models and Languages of Object-Oriented Databases.* Addison-Wesley, Harlow, Engeland.

[5] Paredaens, J. et al. (1989). *The Structure of the Relational Database Model.* EATCS: Monographs on Theoretical Computer Science, Vol. 17, Springer-Verlag, Berlin, Heidelberg.

[6] Zadeh, L. A. (1997). *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic.* Fuzzy Sets and Systems, Vol. 90, No. 2, pp. 111–127.

Management of an estate agency allowing fuzzy data and flexible queries

José Galindo[1] Juan M. Medina[2] Juan C. Cubero[2] Olga Pons[2]

[1]Dpto. Lenguajes y Ciencias de la Computación, Universidad de Málaga (Spain). ppgg@lcc.uma.es

[2]Dpto. Ciencias de la Computación e I.A., Universidad de Granada (Spain). {medina,carlos,opc}@decsai.ugr.es

Abstract

As main result of all our previous work, we have now available a FSQL Server *for Oracle Databases, programmed in* PL/SQL. *This server allows us to query either a Fuzzy or Classic Relational Database with the* FSQL *language (Fuzzy* SQL*). The* FSQL *language is an extension of the* SQL *language which permits us to write flexible (or fuzzy) conditions in our queries to a fuzzy or traditional database. In this work we present a management system for a real estate agency in which some attributes of the landed properties may be fuzzy, i.e., we can store imprecise information about it. Besides, our system allows the user to make flexible or fuzzy queries in order to retrieve the most relevant properties of our database, starting with the customer information. The goal is to retrieve the most interesting landed properties according to the initial customer preferences. Of course, we can obtain a membership degree for each landed property in the fuzzy query result.*

KEYWORDS: Fuzzy Management, Information Retrieval, Flexible Queries, Fuzzy SQL, Fuzzy Query Languages, Fuzzy Relational Databases.

1 Introduction

In the last years, the management applications have been popularized and widely extended. In these applications, the database management is very important. Thus, many times the DBMS (Database Management Systems) are used by the final users directly or through an interface program (front-end).

On the other hand, the fuzzy databases have been developed in the last years, rising up different models, among which they highlight the Prade-Testemale model [8], the Umano-Fukami model [9], the Buckles-Petry model [1], the Zemankova-Kaendel model [10] and the GEFRED model by Medina-Pons-Vila [7]. This last model represents an eclectic synthesis of the different models which have appeared to deal with the problem of the representation and management of fuzzy information in relational databases. One of the main advantages of this model is that it consists of a general abstraction which allows us to deal with different approaches, even when these may seem disparate.

Besides, for the GEFRED model, the FSQL language, a fuzzy (or flexible) query language based on the SQL language, has been defined. This language allows us to express sentences taking into account the characteristics of imprecise information. Thus, as we will see, the FSQL queries allow to express fuzzy conditions, to calculate fulfillment degrees, to establish fulfillment thresholds...

In this paper we intend to discuss about how to use FSQL and a FRDB (Fuzzy Relational DataBase) for the management of a company. We will study the case of an estate agency, devoted to the sale and rent of apartments, flats, houses, chalets, semi-detached houses, building sites, coach houses, industrial plants... Other applications for our FSQL Server can be found, for example, in [2, 5].

First, we include a brief explanation of the main advantages of the FSQL SELECT sentence, in order to express fuzzy queries (a more detailed description of this and other FSQL sentences can be found in [3, 5]). Next, we will expose briefly the information that is stored in the database (fuzzy attributes types...). Finally, we will study how a FRDB can improve the information management system of an estate agency.

2 Flexible Queries with FSQL

The FSQL language [3, 4, 5] extends the SQL language to allow flexible queries. We have extended the SELECT command to express flexible queries and, due to its complex format, we only show here an abstract with the main extensions added to this command:

- **Linguistic Labels**: If an attribute is capable of fuzzy treatment then linguistic labels can be defined on it. These labels will be preceded with the symbol $ to be easily distinguished. Every label has an associated trapezoidal possibility distribution (Figure 1) or there is a similarity relationship defined between each two labels in the same domain.

fulfillment threshold and, if the query retrieves too few properties then we can use possibility comparators instead of necessity ones, which are more restrictive.

Another possibility is to consult *on-line* the FRDB at the same time as the customer indicates his preferences. For example, let us suppose that a customer indicates: "I am looking for a big chalet with about 7 rooms and in the Northern area". Then, the following FSQL query retrieves the properties that comply with those conditions, ranking by the first attribute, which is the compatibility degree:

```
SELECT CDEG(*), Sales.* FROM Sales
WHERE Kind     FEQ  $Chalet .5
AND Surface    FGEQ $Big    .5
AND Rooms      FGEQ #7      .5
AND District FEQ  $North  .5 ORDER BY 1 DESC;
```

We have opted to use possibility comparators because quite a lot of elemental conditions are included, in order to select a greater quantity of properties. In the previous query, the semi-detached houses would also be retrieved if this Kind of property has a similarity degree greater or equal to 0.5 with regard to chalet. If we look exclusively for chalets we must establish the threshold to 1.

It is easy to note that the number of possible queries and the utility of its answers is tremendous. Thus, we will show first to each customer the property that has a greater compatibility degree. In the case that none of the retrieved properties satisfies the customer, we can make a more flexible query, putting down the thresholds (until 0), changing the fuzzy constants on the right of the simple conditions, changing fuzzy comparators, eliminating not very important conditions or exchanging some logical comparator AND by OR.

Naturally, the flexible query system does not assures the accomplishment of operations, but it assures that we find the property most accordant with the customer needs and tastes. It is necessary to take into account that when somebody looks for any type of property, he rarely has a fixed idea, but looks for something with some initial basic characteristics. It is frequent that what the customer acquires finally is not very similar to what he looked for at first.

Besides, this system allows the real estate agency to maintain a large database without having to remember the characteristics of the properties. This problem makes impossible to handle many properties effectively, and it is solved by our proposed system.

5 Conclusions

We have presented a discussion in which it remains clear how the fuzzy databases systems can provide a lot of advantages to the enterprises that use traditional databases. This is possible because we have a flexible query language, the FSQL language, which is very similar to SQL and therefore it is easy to learn and use. The FSQL language has a very similar syntax to standard SQL and, furthermore, it is easily installable on a system which uses Oracle as a DBMS. Of course, the FSQL queries may be implemented in a program in such a way that the user does not need to know the FSQL syntax. Furthermore, we are now working on the creation of a Visual FSQL Client program for Internet (in Java). Others applications for FSQL are shown in [2, 5, 6].

The power of this language, the great quantity of implemented fuzzy comparators, the flexibility to establish fulfillment thresholds, the possibility to be installed and used on traditional DBMS and other shown characteristics make that the advantages of FRDB are easily evaluated and therefore, this means an important step in the transfer of results of research to the business world.

References

[1] B.P. Buckles, F.E. Petry, "Extending the Fuzzy Database with Fuzzy Numbers". Information Sciences 34, pp. 45–55, 1984.

[2] R.A. Carrasco, J. Galindo, M.A. Vila, J.M. Medina, "Clustering and Fuzzy Classification in a Financial Data Mining Environment". 3th International ICSC Symposium on Soft Computing, SOCO'99, pp. 713–720, Genova (Italy), June 1999.

[3] J. Galindo, J.M. Medina, O. Pons, J.C. Cubero, "A Server for Fuzzy SQL Queries", in "Flexible Query Answering Systems", eds. T. Andreasen, H. Christiansen and H.L. Larsen, Lecture Notes in A.I. (LNAI) 1495, pp. 164–174. Ed. Springer, 1998.

[4] J. Galindo, J.M. Medina, A. Vila, J.C. Cubero, "Fuzzy Comparators for Flexible Queries to Databases". Iberoamerican Conference on Artificial Intelligence, IBERAMIA'98, Lisbon (Portugal), 1998.

[5] J. Galindo, "Tratamiento de la Imprecisión en Bases de Datos Relacionales: Extensión del Modelo y Adaptación de los SGBD Actuales". Ph. Doctoral Thesis, University of Granada (Spain), March 1999.

[6] J. Galindo, M.C. Aranda, "Gestión de una Agencia de Viajes usando Bases de Datos Difusas y FSQL". Turismo y tecnologías de la información, TuriTec'99, Málaga (Spain), September 1999.

[7] J.M. Medina, O. Pons, M.A. Vila, "GEFRED. A Generalized Model of Fuzzy Relational Data Bases". Information Sciences, 76(1–2), pp. 87–109, 1994.

[8] H. Prade, C. Testemale, "Fuzzy Relational Databases: Representational issues and Reduction Using Similarity Measures". J. Am. Soc. Information Sciences 38(2), pp. 118–126, 1987.

[9] M. Umano, S. Fukami, "Fuzzy Relational Algebra for Possibility-Distribution-Fuzzy-Relational Model of Fuzzy Data". Journal of Intelligent Inform. Systems, 3, pp. 7–28, 1994.

[10] M. Zemankova-Leech, A. Kandel, "Implementing Imprecision in Information Systems". Information Sciences, 37, pp. 107–141, 1985.

TYPICALITY BASED ON SOFT AGGREGATIONS IN FUZZY OBJECT ORIENTED DATABASES

Gloria Bordogna and Gabriella Pasi
Istituto per le Tecnologie Informatiche Multimediali
CNR - Via Ampere 65, 20131 Milano, Italy
Tel. +39-0270643257/8 Fax: +39-0270643292
Email: fuzzir@itim.mi.cnr.it

Abstract

In this paper a method is proposed to compute typical objects for the classes in the scheme of a fuzzy object oriented database. Typical objects are considered as "representatives" of a fuzzy majority of the class instances. The instances of a class are represented in a topological space and the typical object is derived as "closest" to the fuzzy majority of the class instances.

1. INTRODUCTION

In the conceptual scheme of an Object Oriented Data Base a class identifies the structure shared by a collection of objects; the intensional definition of a class is the specification of its structure in terms of a collection of attributes and methods to operate on them. The extensional definition of a class is the set of its instances, i.e. objects corresponding to real entities of the considered application [1].
The concept of typicality in the OO data model has been introduced to characterize the typical attribute values of a class, thus defining a "virtual-object" which constitutes the class representative.
The primary use of typical objects is to support the association of default attribute values when the actual values are unknown. The use of typical objects makes the data model more informative than when using null values as pointers to unknown attribute values [2, 6, 10] . Moreover typical objects can be useful in database summarization, as they constitute a synthetic view of database contents [12]. Last but not least, in Fuzzy Object Oriented data models typical attribute values can be used to support the computation of the partial membership of objects to fuzzy classes [8].
Generally typical attribute values are defined a priori, in the phase of the data scheme generation, and stored as components of the intensional definition of classes: a limitation of this approach is that the objects stored successively as instances of a class may have attributes which are very different from those previously identified as the typical ones.
In this paper we propose a method to compute typical objects "a posteriori" of the database instantiation. We consider a fuzzy OO data model, the instances of which may have either precise or vague values, defined as trapezoidal possibility distributions over the attribute domain [4,5]. The proposed method is defined as an extension of a method previously defined to compute typical objects for crisp OO data bases [3].
In the literature, it has been outlined that the concept of typicality has a vague nature [7]; in our approach vagueness is modelled by defining vague typical attribute values as well as by computing the typical object as a representative of a fuzzy majority of the class instances. The concept of a fuzzy majority expressed by a linguistic quantifier such as *most of*, was first introduced in the context of fuzzy group decision making to identify a sub-group of consensual experts [9]; in our context this concept is inherited to identify a subset of homogeneous class instances.
In section 2 we synthetically describe the topological model to compute typical objects; in section 3 the representation of the class instances in the topological space are described and in section 4 the computation of typical objects is defined.

2. OVERVIEW OF THE PROPOSED APPROACH

The procedure for the computation of the typical object is based on the definition of a topological space. In the

crisp environment, for each class in the database scheme a D-dimensional space is defined, in which the instances are represented as points whose coordinate values are the precise attribute values [3]. The typical object is computed as a new point obtained by the aggregation of the instance points through a linguistic quantifier; the aggregation is performed separately on each coordinate (referring to a distinct attribute) of the points.
In the fuzzy extension proposed in this paper we consider that the attribute values may be vague, identified by linguistic labels associated with possibility distributions [4,5].
Let us try to figure out how an instance of a class looks like in the space. For sake of simplicity, let us first restrict our attention to a three dimensional space and to a class instance for which two attribute values are known precisely a_1, and a_2, while the third attribute has a vague value represented by a possibility distribution π_3. Now, let us assume that the possibility degrees are associated with grey levels; the value 1 is associated with the black, the value 0 with the white. In the space, the class instance identifies a segment shaded with different grey levels (see Figure 1).

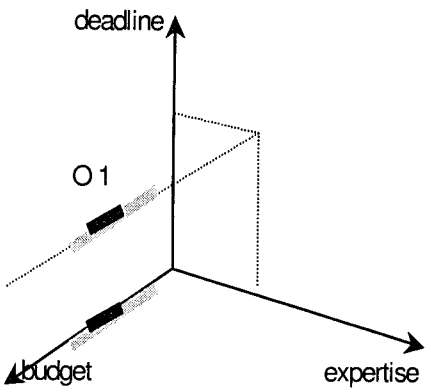


Figure 1: representation of a class instance.

By generalizing this view to the extreme case in which all the attribute values are vague, a class instance identifies an ipercube with a black nucleous and surrounding shells shaded with different grey levels. In this general case, the grey level associated with a point of the ipercube is determined by a combination of the grey levels of its coordinate points. In the following, we will figure out a class instance as an ipercube in the space, even if this happen only as extreme case.
This graphical representation of the class instances is intuitive and makes it possible to identify the typical object of a class as the "virtual" ipercube closest to the fuzzy majority of the class ipercubes. It is determined by computing the possibility distributions characterizing the typical vague values of the attributes for each coordinate axis; this computation consists in the aggregation of the trapezoidal functions representing the attribute values of the class instances through an Induced Ordered Weighed Averaging operator reflecting the fuzzy majority [14]. This computation is based on the adoption of both a metrics for each coordinate of the space and a distance function in the space.
Further, a degree of consensus or agreement on the typical object among the instances in the fuzzy majority is evaluated: it is interpreted as an indicator of the *cohesion* of the class instances in the fuzzy majority. When the cohesion degree is low, i.e. below a given threshold, it means that the typical object is not a valid representative of the instances; in this case a new typical object representative of a stricter fuzzy majority of the instances can be computed.

3. A TOPOLOGICAL REPRESENTATION OF OBJECT INSTANCES

In database applications the available information is often characterized by *incompleteness*: a datum may be either ill-known (vague or uncertain), or completely unknown or it may not exist. Usually in databases when there is a lack of knowledge about the value of an attribute a *null value* is used as a place holder for the unknown value. Different kinds of null values have been introduced in the literature to characterize the different types of incompleteness [2,6,18].

In OOD models the computation of typical attribute values as prototypal values of the attributes of each class would offer a more infomative way than null values to represent unknown information.
Recently fuzzy set and possibility theory have been applied to provide a more flexible and accurate way to model incompleteness in data models. To this aim, several fuzzy Object Oriented data models have been proposed in the literature [4, 8, 11, 15]. Some of these models assume that at the schema level, typical attribute values are defined [8, 10].
We call typical object a representative instance of a structured class (a class having more attributes), while a representative instance of an attribute is a typical attribute value. A typical object is characterized by a set of typical attribute values.
The specification of typical attribute values is usually assumed as an a priori operation performed at the scheme definition level by an expert of the application domain. In [3] a method has been proposed to compute typical objects "a posteriori" of the database instantiation as representatives of a fuzzy majority of the class instances. To this aim, for each class in the scheme a topological space is defined with a dimension D given by:

$$D = \sum_{i=1}^{N} D_i$$

where:

- N is the number of the attributes of the class which are semantically significant in defining the typical object: for example the attributes having a numeric domain, such as the attribute age, or an ordinal domain, such as the attribute hear color. In fact, it makes sense to define a typical age of a population or a typical hear color, while it is a nonsense to define the typical name.
- D_i is the dimension of the space associated with the i-th attribute of the class. $D_i = 1$ for the attributes having a primitive domain. In this context only single valued attributes are considered. Further it is not allowed to choose as semantically significant the class attributes with have the class itself as domain.

In the crisp context, each coordinate of the topological space is associated with an attribute having a primitive domain. The instances of the class identify points whose coordinates are the precise values of the associated attributes.
In the fuzzy extension, both the instances of a class and the typical object are identified by ipercubes of the D-dimensional space, as it has been illustrated in section 2. Each axis of this space is associated with the basic domain of a "significant" attribute of the class; the attribute values of a class instance are defined by possibility distributions with a trapezoidal shape $(\alpha, \beta, \chi, \delta)$ on the attribute basic domain; when an attribute value is not vague (it is either precise or imprecise), we will assume that it is represented by a possibility distribution too (for a precise value $\alpha = \beta = \chi = \delta$; for an imprecise value $\alpha = \beta$ and $\chi = \delta$).

4. TYPICAL OBJECT COMPUTATION

The procedure for the computation of the typical object of a class is the following. First, for each class a subset of its instances is selected randomly as a sample set; it is assumed that the class instances are represented by this sample set. The sample set cardinality (K) strongly influences the time needed for the computation of the typical object.
An evaluation matrix can be defined in which the rows refer to the K instances of the class and the columns refer to the D attributes (i.e. the space dimension). The value $\pi_{ij} = (\alpha_{ij}, \beta_{ij}, \chi_{ij}, \delta_{ij})$ of the matrix is the trapezoidal possibility distribution of the j-th attribute for the i-th instance.

Class C	A1	A2	A3	A4	A5
Object 1					
Object 2		π_{22}			
Object 3				π_{34}	

Table 1: matrix representing the instances of a class

The typical object of a class is computed so as to take into account a fuzzy majority Q, such as "*most*" [9, 13], of the K class instances: it can be also figured out as an ipercube $\underline{o}$ in the D-dimensional space of the attributes closest to the ipercubes in the fuzzy majority Q.
The rows of the decision matrix, are used to compute the typical object $\underline{o}$.
The aggregation function is formalized by an IOWA operator of dimension K, with the weighting vector W automatically defined so as to reflect the semantics of the quantifier Q [13]. First, by following Zadeh [16], the membership function of the fuzzy subset representing the relative monotone non decreasing quantifier is defined $Q : [0,1] \rightarrow [0,1]$. In fact, relative monotone non decreasing quantifiers model the semantics associated with the concept of a fuzzy majority which increases as the number of the elements in the fuzzy majority come close to *all*.
Then the K elements $w_i \in [0,1]$ of the weighting vector W are obtained as:

$$w_i = Q(i/K) - Q((i-1)/K) \quad \forall i=1 \ldots K$$

We remind that an Induced Ordered Weighted Average operator (IOWA) allows to generalize the OWA so as the non linear component requiring the reorder of the arguments can be defined based on a reorder vector R [14].
In our context, since the possibility distributions are represented by trapezoidal fuzzy numbers $(\alpha, \beta, \chi, \delta)$, we apply the IOWA separately to the bags:
$(\alpha_1, \ldots, \alpha_K)$, $(\beta_1, \ldots, \beta_K)$, $(\chi_1, \ldots, \chi_K)$, $(\gamma_1, \ldots, \gamma_K)$, by considering a common reorder vector R.
By considering the aggregation of the $(\alpha_1, \ldots, \alpha_K)$, the IOWA operator is defined as follows:

$$IOWA(\langle r_1, \alpha_{1j} \rangle, \ldots, \langle r_K, \alpha_{K\,j} \rangle) = \sum_{i=1}^{K} b_i w_i$$

in which the value b_i is the α_j that has associated the i-th largest of the r values.
The reorder vector R is common to both the aggregations on all the axis, and to the aggregations of all the four bags along each axis. Its elements are determined on the

basis of the euclidean distance of the instances from their average instance.
The average instance is obtained by applying an average operator to the values (α, β, χ, δ), in each column of the decision matrix.
A distance measure between possibility distributions is adopted : let us indicate by $d_h(\pi_{ih}.\pi_{jh})$ the distance on the h-th axis [17]. Then, the Euclidean distance between two class instances d (o_i o_j) is obtained as:

$$d(o_i\ o_j)=\sqrt{d_1(\pi_{i1}\ .\pi_{j1})^2+\ d_2(\pi_{i2}.\pi_{j2})^2+\ ..\ d_D(\pi_{iD}.\pi_{jD})^2}$$

The K instances are then ranked based on R in decreasing order of their euclidean distance from the average instance distance $d(o_i\ ,\ o_{Average})$.
By using this reorder criterion the typical attribute values are determined more heavily by the instance attribute values which are "closer" to the average value of the attributes. The typical attribute values computed in this way are seen as representatives of Q attribute values, and the typical object $\underline{o}$ is the representative of Q instances.
The degree of cohesion of the class instances in the fuzzy majority with respect to the typical object is computed as follows:

- first, the similarity between each instance o_i and the typical object $\underline{o}$ is evaluated; we interpret this evaluation as the computation of the *degree of cohesion of the instance with respect to Q instances*; the similarity measure between pairs of instances is defined as the complement of their Euclidean distance normalized with respect to the maximum:

$$sim(o_i\ \underline{o}) = 1 - \frac{d(o_i\,\underline{o})}{\max_{j=1,\ldots K}(d(o_j\,\underline{o}))}$$

- second, the *degree of cohesion among Q instances* is computed; this is performed by aggregating the similarity degrees computed in the previous step by using the OWA operator associated with Q.

When the degree of cohesion is not satisfactory, i.e. it is below a given threshold, it is assumed that the typical object is not a valid representative of the considered majority. In this case, one can try to compute a new typical object as representative of a stricter fuzzy majority.

References

1. E. Bertino, and L. Martino, Object-Oriented Database Management Systems: Concepts and Issues. *IEEE Computer* 4, 33 ,1991.
2. J. Biskup, A formal approach to null values in database relations, in *Advances in Database Theory*, H. Gallaire & J.M. Nicolas eds., 1, Plenum Press, 299, 1981.
3. G. Bordogna and G. Pasi, "Typicality based on soft Aggregations in Crisp Object Oriented Database", in proc. *of NAFIPS'99*, New York, 9-11 June, 1999.
4. G.Bordogna and G. Pasi, D. Lucarella, A Fuzzy Object Oriented Data Model for Managing Vague and Uncertain Information, to appear in *International Journal of Intelligent Systems*, 1999.
5. P.Bosc, and H. Prade, An introduction to fuzzy set and possibility theory-based treatment of soft queries and uncertain or imprecise databases. *Report IRIT/93-57-R*, 1993.
6. E. F. Codd, Missing Information (applicable and inapplicable) in relational databases, *SIGMOD record*, 15, 53, 1986.
7. D. Dubois, H. Prade, J-P. Rossazza, Vagueness, Typicality and Uncertainty in Class Hierarchies, *Int. Journal of Intelligent Systems*, 6, 167, 1991.
8. R.George, B.P. Buckles, F.E. Petry, Modelling Class Hierarchies in the Fuzzy Object-Oriented Data Model, *Fuzzy Sets and Systems* 60(3), 1993.
9. J. Kacprzyk, M. Fedrizzi and H. Nurmi, Fuzzy Logic with Linguistic Quantifiers in Group Decision Making and Consensus Formation, in An Introduction to Fuzzy Logic Applications in Intelligent Systems, R.R. Yager and L.A. Zadeh eds., Kluwer, 263-280, 1992.
10. G. Pasi , and R. R. Yager, Calculating Attribute Values using Inherithance Structures in Fuzzy Object Oriented Data Models, to appear in *IEEE Transactions on Systems, man and Cybernetics*, 1999.
11. R. De Caluwe, *Fuzzy and Uncertain Object-Oriented databases, Concepts and Models*, World Scientific, Singapore, 1997.
12. R.R. Yager, A new approach to Summarization of data: Information Science, 28, 69-86, 1982.
13. R.R. Yager, Interpreting Linguistically Quantified Propositions, *International Journal of Intelligent Systems*, 9, 541, 1994.
14. R.R Yager., D. Filev, Operations for Granular Computing: Mixing Words and Numbers, in proc. of *FUZZIEEE*, 123, 1998.
15. N.Van Gyseghem, R.de Caluwe, R. Vandenberghe, UFO: Uncertainty and Fuzziness in an Object-Oriented model, *II IEEE Int. Conf. on Fuzzy Systems*, 1, 489 (San Francisco, USA, 1993.
16. L.A. Zadeh, A computational Approach to Fuzzy Quantifiers in Natural Languages, *Computing and Mathematics with Applications*. 9, 149, 1983.
17. L.A. Zadeh, Similarity relations and fuzzy orderings. Information Science, 3, 177-200, 1971
18. R. Zicari, Incomplete Information in Object-Oriented Databases, *SIGMOD RECORD*, 19, 33-40 1990.

Implementation of an Inference Engine for Fuzzy Databases

I. Blanco iblanco@decsai.ugr.es
J. C. Cubero JC.Cubero@decsai.ugr.es
F. Cuenca fcuenca@decsai.ugr.es
O. Pons opc@decsai.ugr.es

Department of Computer Science and Artificial Intelligence
University of Granada
Avda. Andalucía, 38
18071 Granada (SPAIN)

Abstract

This work shows how a relational DBMS with deductive and imprecise capabilities can be improved by adding the capability of representing and handling imprecise values to an existing classical inference engine that works on an RDBMS. The possibility of making flexible queries on classical values stored in the database is also added.
Keywords: relational databases extension, fuzzy deduction, inference.

1 Introduction

Since the introduction of the relational model by Codd in [1], it has been used in a generalized way by many authors.

In such a model, we can store information about people's ages, and this can be recovered if we want to find out a specific person's age or even an interval containing such a value. But this type of information is not usual when a human being makes a reasoning. We usually ask for "young" or "old" people without considering that such concepts vary from one person to another, and that the edges of such concepts are not well-defined.

In order to accomplish this task, we have to distinguish between how such information is represented and handled. To deal with the problem of the representation of fuzzy information in the relational model, Medina, Pons and Vila presented GEFRED [4] but this is only one part of the problem studied. We also have to model another way of implicit reasoning which a person can use.

When a human being thinks, he obtains data from other data, i.e. he is able to deduce information that depends on itself. For instance, a person's father, mother, and ancestors can be stored in separate tables, but ancestors can be obtained from the father and mother tables and queried in a fixed order. Another problem arises if we store ancestors extensionally, that of the subsequent modification of the ancestor table when a modification of the fathers or mothers table occurs.

When we obtain data from other data, we follow a set of steps or "rules" which we could represent in a logic structure and we could design an algorithm that uses this structure to access the tables to complete the dependent table.

In order to implement both fuzzy and deductive features in a classical RDBMS, Medina, Pons, Vila and Cubero introduced the FREDDI system in [3].

Consequently, we would have a representation of the flexible values and a language for querying them, a flexible query language for querying concrete values, a structure to represent logical rules, and an algorithm to deduce data from other ones.

2 Considerations of implementation

To develop the inference engine, two possibilities have been taken into account:

- *Weak coupling:* this is orientated to host languages, that is, general purpose programming languages that have been extended with database communication features. An inference engine designed using this type of language is a *built-out engine*, in the sense that it is external to the RDBMS and requires a network or some type of message protocol to communicate with the database.

- *Tight coupling:* this is orientated to database programming languages such as language PL/SQL© provided by Oracle©, that are less generic but which are incorporated into the RDBMS. A module developed with such languages is usually executed by the RDBMS. As a result, an inference engine developed with a language of this type is a *built-in inference engine.*

Once we have examined the two possibilities, we shall have to consider portability versus

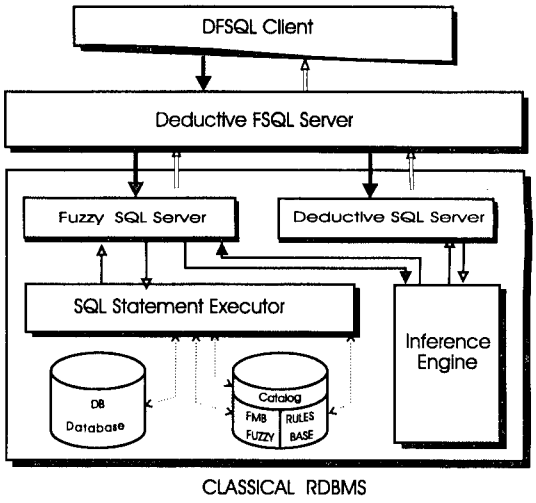


Figura 1: Architecture for Fuzzy Deductive SQL Server

efficiency. Due to the amount of data handled by a system with deductive capabilities, the Built-in Inference Engine is more efficient for our purposes because database communication capability is not required and the RDBMS executes the module more efficiently than any external module. The architecture of the system is shown in figure 1. Once our choice has been made, we describe it taking into account the problems of both representation and handling associated with deductive and fuzzy data.

3 Classical inference engine

As introduced in [3], two types of tables can be stored in a database with deductive capabilities. The first type are the so-called *extensional tables* whose content is stored explicitly in the table. The second type are the so-called *intensional tables* whose content depends on the contents of other tables.

Let us imagine a table representing a relation, for instance, a table called PARENT that

stores the parent together with each son. Now, let us suppose that we want to retrieve and store information about ancestors in a table whose content depends on the PARENT table. We could store this information explicitly but every time the PARENT table changes, we would have to manually change the ANCESTOR table. We could make ANCESTOR in some way dependent on PARENT. If we were to do this, then the process of updating it can be made automatically. The PARENT table can be seen as a predicate PARENT (X, Y) and each tuple stored in the table as a fact, for instance, PARENT ('Mike', 'Susan'). From this point of view, we could associate extensional tables to a type of predicate that we shall call "extensional" and intensional tables to "intensional" predicates. Once the tables have been represented logically, we can describe the relations established between intensional and extensional tables by using logical rules. An intensional predicate is described as a disjunction of sub-rules containing intensional or extensional predicates that are joined by a conjunction operator.

At this point, a query to the database can involve both extensional or intensional tables. The sub-query involving an extensional table is solved classically, that is, with a classical SELECT sentence. But the sub-query involving an intensional one is solved by filling in the table using the logical rules that define the condition that tuples satisfying such a predicate have to verify.

The algorithm to expand the rules that define an intensional predicate is based on the backtracking mechanism and on the resolution inference algorithm used to do this task, that is, variables instantiation. Each rule starts with a set of free variables on which predicates that are part of the rule will be applied, and some or all of the variables will be fixed with a value. When the algorithm applies a predicate to the set of variables associated to a rule, it can obtain more than one combination of values so at this point we shall have a set of several variables instead of one which is no longer valid because it has been processed. Once the algorithm has processed the current set, it can be deleted and another one becomes the active set. When the algorithm finishes, we shall have sets of several variables that are fully, partially or not instantiated. The set of fully instantiated variables has predicate results because a combination of values satisfying the rule has been found. Those that are partially or not instantiated are ignored because such a combination of values has not been found.

4 Fuzzy extension of the deductive module

Once we have studied how to improve a classical RDBMS with deductive capabilities, we accomplish the problem of incorporating the handling (representation and deduction) of fuzzy information. To achieve this, we start from the theoretical model for representing fuzzy information in the relational model presented in [4] and the concrete implementation of this introduced by J. Galindo in [2]. On the other hand, we use the theoretical model to represent deductive information presented in [3].

When we apply predicates of a rule to a variables set, we operate as follows: when the first predicate is applied, no variable is instantiated so the values combination returned corresponds to all tuples contained in the table

associated to the predicate. But from this moment, some variables are instantiated in all variables set. This is not a problem in classical deduction because, in this case, when the next predicate is applied only those tuples whose values are equal to non free variables values are important. In classical deduction, the comparator used is equality but now fuzzy comparators must be considered as a coincidence or matching measure.

But this is not the only consideration to be taken into account. Once we have decided which fuzzy comparator has to be applied, it cannot be applied directly. In classical deduction, only those tuples whose values coincided with non free variables values appeared in the result. But in fuzzy terms, all values in a domain coincide with a concrete value with a degree that can be zero. Exploring all values can increase the number of variables sets exponentially and some of them can be irrelevant. So we have to modify the model in such a way that allows us to associate a threshold to each predicate in a rule if necessary, or even to each argument contained in these predicates. Finally, DDL sentences will have to be modified in order to accept this type of fuzzy logical rules.

5 Conclusions

This work shows how the relational model can be improved and accepts the representation and management of fuzzy and/or deductive information. This part is an extension of works that separately manage fuzzy information or deductive information.

Advantages and handicaps of the two possibilities for the inference engine implementation are analyzed. Finally, a built-in solution is chosen since it offers more advantages in data transfer between the DBMS and the inference engine. If we consider the implementation assumed, the language SQL has been improved so that it can accept the new data types introduced (DDL) that can be fuzzy or intensional, and manipulate such data types (DML) (so far, only relevant sentences have been implemented).

In the future, we will add an importance degree to the rules so that tuples obtained by expansion of one or other rule do not have the same weight. In order to perform this task, we will have to unify the imprecision measures obtained with the importance value associated to the rule.

References

[1] E.F. Codd. A relational model of data for large shared data banks. *Communications of the ACM*, 13(6):377–387, 1970.

[2] J. Galindo. *Tratamiento de la Imprecisión en Bases de Datos Relacionales: Extensión del modelo y adaptación de los SGBD actuales.* PhD thesis, Department of Computer Sciences and Artificial Intelligence, University of Granada, SPAIN, 1999.

[3] J.M. Medina, O. Pons, J.C. Cubero, and M.A. Vila. Freddi: A fuzzy relational deductive database interface. *International Journal of Intelligent Systems*, pages 597–613, 1997.

[4] J.M. Medina, O. Pons, and M.A. Vila. Gefred: A generalized model of fuzzy relational databases. *Information Sciences*, 76(1-2):87–109, 1994.

Fuzzy Clustering with Weighting of Data Variables *

Annette Keller
Institute for Flight Guidance
German Aerospace Center
Lilienthalplatz 7
D-38108 Braunschweig, Germany
e-mail annette.keller@dlr.de

Frank Klawonn
Dept. of Electrical Engineering and Computer Science
Ostfriesland University of Applied Sciences
Constantiaplatz 4
D-26723 Emden, Germany
e-mail klawonn@et-inf.fho-emden.de

Summary

We introduce an objective function-based fuzzy clustering technique that assigns one influence parameter to each single data variable for each cluster. Our method is not only suited to detect structures or groups in unevenly over the structure's single domains distributed data, but gives also information about the influence of individual variables on the detected groups. In addition, our approach can be seen as a generalization of the well-known fuzzy c-means clustering algorithm.

Keywords: Fuzzy clustering, variable selection, generalized fuzzy c-means

1 Introduction

The common objective function to be minimized in fuzzy clustering is of the form

$$J(X,U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m d^2(v_i,x_k) \qquad (1)$$

where c is the number of fuzzy clusters, $u_{ik} \in [0,1]$ is the membership degree of datum x_k to cluster i and $d(v_i,x_k)$ is the distance between cluster prototype v_i and datum x_k. In order to avoid the trivial solution $u_{ik} = 0$, additional assumptions have to be made leading to probabilistic [1], possibilistic [8] or noise [2] clustering. Parameter $m \in \mathbb{R}_{>1}$ is called fuzzifier. For $m \to 1$, we have for the membership degrees $u_{ik} \to 0/1$, so the classification tends to be crisp. If $m \to \infty$, then $u_{ik} \to \frac{1}{c}$, where c is the number of clusters.

* This work was supported by the European Union under grant EFRE 98.053

The prototypes can be simple vectors like the data as in the fuzzy c-means algorithm (FCM) or more complex structures like in the Gustafson-Kessel algorithm [4], in linear or shell clustering [7]. In these cases the distance function d is not simply the Euclidean distance but some other measure depending on the type or form of the clusters. A thorough overview on objective function-based fuzzy clustering techniques can be found in [5].

In this paper we introduce a new distance measure that generalizes the FCM in adding a parameter that determines the influence of certain data attributes for some cluster.

2 Attribute Weighting Fuzzy Clustering

Especially in data where few variables determine particular clusters other variables may disguise the structure and should therefore not be considered to find these clusters. This can be done by weighting single attributes for each cluster as we have done with our new distance measure.

The distance between a datum x_k and a cluster (vector) v_i is defined by

$$d^2(v_i,x_k) = \sum_{s=1}^{p} \alpha_{is}^t \cdot \left(x_k^{(s)} - v_i^{(s)}\right)^2. \qquad (2)$$

$x_k^{(s)}$ and $v_i^{(s)}$ indicate the sth coordinates of the vectors x_k and v_i, respectively. The number of variables or attributes is denoted by p. α_{is}^t is a parameter determining the influence of attribute (coordinate) s for cluster i. The parameters α_{is} can be considered as fixed or adapted individually for each cluster during clustering due to the constraint

$$\forall i \in \{1,\cdots,c\}: \quad \sum_{s=1}^{p} \alpha_{is} = a, \qquad (3)$$

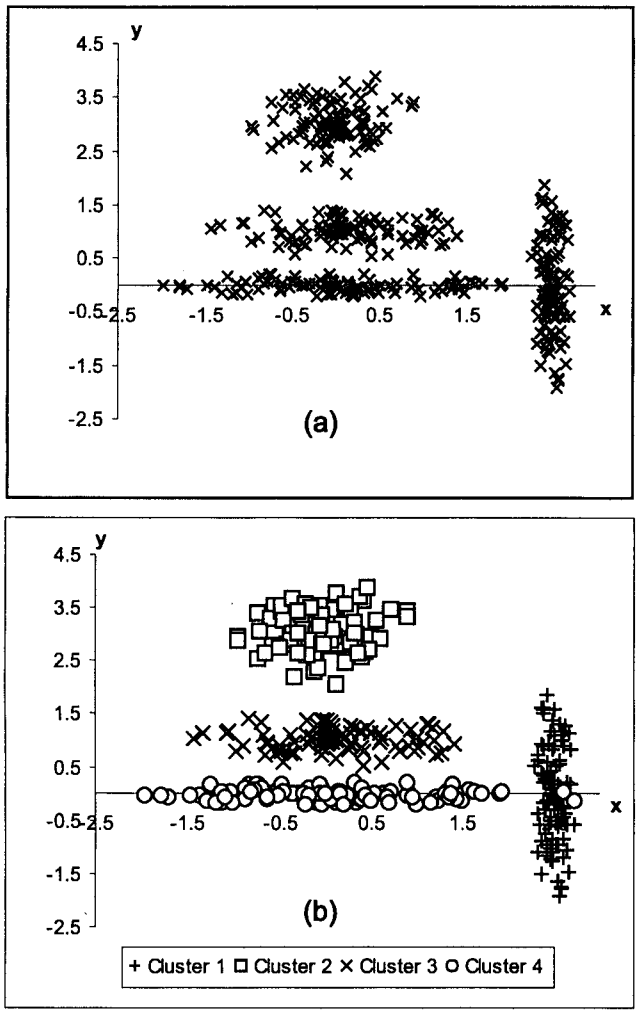


Figure 1: Ellipsoidal clusters

where $a \in \mathbb{R}$ is a constant parameter, e.g. $a = 1$ or $a = c$. If we would neglect this constraint, we would obtain the trivial solution $\alpha_{is} = 0$ for all i and s.

We will see that the exponent $t \in \mathbb{R}_{\geq 1}$ in equation (2) has a similar influence on the parameters α_{is} as the fuzzifier m on the membership degrees u_{ik}. For $t \to 1$ the α_{is} tend to be 1 or 0 – either one attribute has unrestricted influence or no influence at all. On the other hand, if $t \to \infty$, all attributes get the same influence on the cluster structure, i.e. $\alpha_{is} \to \frac{1}{a}$ for all i and s.

Based on this approach we can derive an alternating optimization scheme for fuzzy clustering using distance measure (2).

To adapt the influence parameters α_{is} we have to determine a necessary condition for the values α_{is} so that the objective function achieves an optimum value.

With condition (3) we obtain the Lagrange function

$$J_\lambda(X,U,v) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m \cdot \sum_{s=1}^{p}\alpha_{is}^t \left(x_k^{(s)} - v_i^{(s)}\right)^2 - \sum_{i=1}^{c}\lambda_i\left(\sum_{s=1}^{p}\alpha_{is} - 1\right). \quad (4)$$

Differentiating (4) leads to equation (5) for the parameter α_{is} as a necessary condition for the objective function to have a minimum. This equation can be used for updating α_{is} during the alternating clustering procedure.

$$\alpha_{is} = \frac{1}{\sum_{r=1}^{p}\left(\frac{\sum_{k=1}^{n} u_{ik}^m \cdot \left(x_k^{(s)} - v_i^{(s)}\right)^2}{\sum_{k=1}^{n} u_{ik}^m \cdot \left(x_k^{(r)} - v_i^{(r)}\right)^2}\right)^{\frac{1}{t-1}}}. \quad (5)$$

In a similar way we obtain a necessary condition for the cluster centres (6).

$$v_i^{(s)} = \frac{\sum_{k=1}^{n} u_{ik}^m \cdot x_k^{(s)}}{\sum_{k=1}^{n} u_{ik}^m} \quad (6)$$

Equation (6) is the same as in FCM.

3 Examples

In Figure 1 a data set consisting of four ellipsoidal groups is shown. Part (a) presents the original data set and part (b) represents the clustering result obtained by our attribute weighting clustering technique where a datum is assigned to the cluster to which it has the highest membership degree (maximum defuzzification). In this case we have set the fuzzifier m and the exponent t to 2.0. For parameter a we have chosen the value 1. However, the clustering result depends more on a suitable initialization of cluster centres than the choice of parameters m, t and a. Table 1 lists the minimum and maximum attribute values for all clusters.

Table 1: Minimum/maximum data values for each cluster

	attributes			
	x		y	
	min	max	min	max
cluster 1	2.28	2.71	−1.93	1.85
cluster 2	−0.96	0.90	2.07	3.88
cluster 3	−1.42	1.40	0.54	1.40
cluster 4	−1.97	1.93	−0.21	0.21

In table 2 *cluster* 1 represents the ellipsoidal group with greatest x-values in the right part of figure 1.

From top to bottom in the left part of figure 1 are the clusters *cluster* 2, *cluster* 3 and *cluster* 4. The scale values α_{is} were adapted during the clustering procedure. It is obvious, that for each cluster the more the data coordinates are scattered around the corresponding prototype's coordinate, the less is the influence of the corresponding attribute for that cluster. In our example in figure 1 the two attribute influence parameters α_{is} for *cluster* 2 have nearly the same value. The data coordinates are approximately uniformly distributed for the two domains of this cluster. For clusters 3 and 4, the data values for attribute x are scattered widely whereas the values for attribute y have a small range – so the influence parameters α_{ix} are small in comparison to α_{iy} for clusters 3 and 4. In case of *cluster* 1 the data values for attribute y are scattered widely, resulting in a high value for influence parameter α_{1x}.

Table 2: Attribute weights for ellipsoidal data set

	attributes	
cluster i	α_{ix}	α_{iy}
cluster 1	0.99	0.01
cluster 2	0.49	0.51
cluster 3	0.08	0.92
cluster 4	0.01	0.99

Figure 2 presents the clustering result for the example data set generated by the FCM clustering technique with fuzzifier $m = 2$ as in our approach. Using the Euclidean distance measure, the FCM is not well suited to detect ellipsoidal structures in data. One indication for the suitability of a clustering result is following value. Of the c membership degrees associated with each datum, we only consider the highest membership degree (i.e. the membership degree to the cluster to which we would assign the datum by maximum defuzzification) and compute the mean value of these membership degrees. Here, the mean value for FCM is 0.81 in comparison to 0.96 for our attribute weighting clustering technique. Nevertheless are the methods from Gustafson and Kessel [4] or Gath and Geva [3] well suited to detect the structures of our example data.

4 Conclusions

The presented clustering technique gives us information about the influence of particular variables or attributes of the data set on special clusters. This knowledge can be used e.g. in classification tasks to determine or detect class defining attributes. Without ignoring one data attribute for the whole classification it is possible to reduce the influence of that attribute on only some clusters. In that way, attribute weights could help to partition the whole data set into smaller data parts depending on the same attributes. Analysing the smaller parts with a reduced number of attributes would reduce the computation effort. Real data sets soon get immense large as e.g. in the mentioned EU project where we analyse sound measured on tyres with different pressures (2520 data sets with 200 sound attributes and 12 different pressures as classification attribute). Here attribute weighting could not only be helpful in reducing computation time but also to reduce the future expense of measuring.

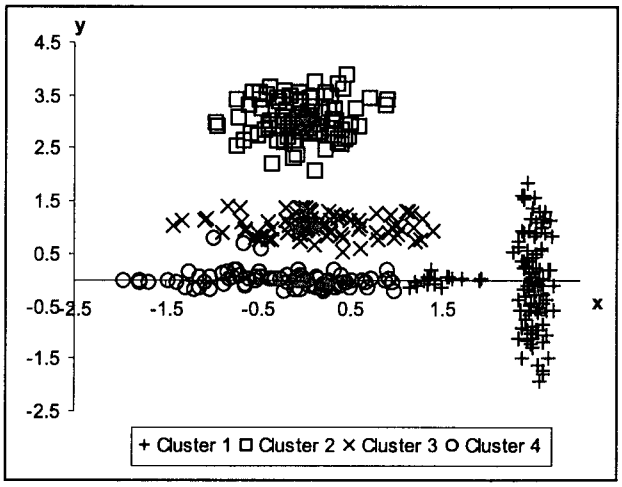


Figure 2: Results for ellipsoidal clusters with Fuzzy-c-means

Our fuzzy clustering approach is also well suited for deriving rules from the clusters. Since the weighting of the attributes for each cluster provides information about the importance of the variables, we can neglect variables with very small weighting factors in the rules. In the example in the previous section this would mean that we would derive a fuzzy rule from cluster 2 invoking only the variable y.

It should be noted that our approach is also related to the simplified version of the Gustafson-Kessel algorithm described in [6] that introduces a diagonal matrix for each cluster. The diagonal elements are weights for the attributes in the same way as we use them here, except for our exponent t. However, the constraint is that the determinant is constant, i.e. the sum in equation (3) is replaced by a product. The advantage of our new approach is that we can control how strong the influence of single variables can be by the parameter t.

Note that our approach differs from the idea to carry out a cluster analysis first and then apply something like a principal component analysis to each cluster.

This would mean that the clustering must take all attributes into account, whereas in our approach the selection of relevant variables is already carried out during the clustering.

References

[1] J. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York, 1981.

[2] R. Davé. Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12:657–664, 1991.

[3] I. Gath and A. Geva. Unsupervised optimal fuzzy clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 11:773–781, 1989.

[4] D. Gustafson and W. Kessel. Fuzzy clustering with a fuzzy covariance matrix. In *IEEE CDC*, pages 761–766, San Diego, 1979.

[5] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler. *Fuzzy Cluster Analysis.* Wiley, Chichester, 1999.

[6] F. Klawonn and R. Kruse. Constructing a fuzzy controller from data. *Fuzzy Sets and Systems*, 85:177–193, 1997.

[7] R. Krishnapuram, H. Frigui, and O. Nasraoui. Fuzzy and possibilistic shell clustering algorithms and their application to boundary detection and surface approximation - part 1 and 2. *IEEE Trans. on Fuzzy Systems*, 3:29–60, 1995.

[8] R. Krishnapuram and J. Keller. A possibilistic approach to clustering. *IEEE Trans. on Fuzzy Systems*, 1:98–110, 1993.

CLUSTER VALIDITY FOR FCM CLUSTERING ALGORITHM USING UNIFORM DATA

Sergio López
Telefónica I+D
Emilio Vargas 6, 28043, Madrid, (SPAIN)
slg@tid.es

Luis Magdalena **Juan R. Velasco**
ETSI Telecomunicación (UPM)
Ciudad Universitaria s/n. 28080, Madrid , (SPAIN)
llayos@mat.upm.es juanra@gsi.dit.upm.es

Abstract

One of the main drawbacks of the FCM clustering algorithm is that it does not calculate the suitable number of clusters. This paper presents a method to solve this problem, by means of an equalization function (using uniform data) for the FCM functional *J*. The results for 2 and 3 dimensional data tests are also presented.

Keywords: Clustering, fuzzy logic, uniform data

1 INTRODUCTION

A clustering algorithm is a mathematical tool that detects similarities in a collection of data. One of the most widely used in different fields, such as pattern recognition, data analysis and image processing is the fuzzy c-means [1] clustering algorithm.

However, FCM has several drawbacks, such as the dependence of the results on the initialization and the need to predefine the number of clusters to be generated. In [2] the authors have presented a genetic fuzzy c-means (GFCM) clustering algorithm, that solves the initialization problem using genetic algorithms. In addition to this, [2] introduced a way of finding the suitable number of clusters for an unidimensional problem. In this paper the authors will deal with the problem of finding the number of clusters for higher dimensions.

2 STANDARD FCM CLUSTERING ALGORITHM

Consider a set of n objects $X = \{x_1, x_2,...,x_n\}$ where $x_i \in R^S$. Each x_i is an object that is described by s real-valued measurements of their features. A fuzzy c-partition of X is a class of c fuzzy sets V_1, V_2, ..., V_C, where c is an integer in the range [2,n]. Then a fuzzy c-partition space for X is the set:

$$M_{fcn} = \left\{ U \in R^{cxn} \middle| u_{ik} \in [0,1], \forall i,k; \sum_{i=1}^{c} u_{ik} = 1, \forall k; 0 < \sum_{k=1}^{n} u_{ik} < n, \forall i \right\} \quad (1)$$

The aim of the algorithm is to find the best partition matrix U in M_{fcn}. This objective is reached when the following function is minimized:

$$J_M(U,V) = \sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}^m \cdot d_{ik}^2(v_i, x_k), \quad U \in M_{fcn}, 1 < m < \infty \quad (2)$$

In this function, v_i are the prototypes (or cluster centroids) of each class, m is a weighting exponent and d is the Euclidean distance. The cluster centroids are obtained with the following expression:

$$v_i = \frac{\sum_{k=1}^{n} u_{ik}^m \cdot x_k}{\sum_{k=1}^{n} u_{ik}^m} \quad (3)$$

The steps of FCM algorithm are the following:

1. Select an initial partition $U^{(0)}$.
2. Compute the cluster centroids v_i.
3. Update the fuzzy membership

$$u_{ik} = \frac{\left(\frac{1}{d_{ik}^2(v_i, x_k)}\right)^{1/(m-1)}}{\sum_{j=1}^{c}\left(\frac{1}{d_{jk}^2(v_j, x_k)}\right)^{1/(m-1)}} \quad (4)$$

4. Repeat steps (2) and (3) until the value of J_M is no longer decreasing.

3 CONCEPT OF CLUSTER SPECIES

One of the main drawbacks of the FCM clustering algorithm is that the number of clusters c is fixed in the beginning of the algorithm. It can be assumed in several problems, where the number of clusters is known. However, there are many problems where is necessary to find the number of clusters.

In [5] the authors have presented the idea of cluster species, based on natural genetics. Given a certain set of solutions to a clustering problem, a cluster species is composed of the solutions having the same number of clusters.

The last step of the process is to apply a quality function to the better individual of each species. This function determines how strong each species is. It will be presented in the next section.

4 QUALITY FUNCTION FOR CLUSTER SPECIES

In FCM, the value of the functional J_M determines how good a clustering partition is. However, due to the definition of J_M, this value decreases as the number of clusters increases.

The solution to this problem is to find a reference case, where there are not any substructures in data. In this way, as the results for any number of clusters must be equivalent, it can be found an equalization function g such as:

$$g(c) \cdot J_m^c(U,V) = K \tag{5}$$

In [2] it was found that the uniform distribution of data is the solution for the unidimensional case. The equalization function g, considering the case of $m=1$, is

$$g(c) = c^2 \tag{6}$$

In the next subsections the 2 and 3-dimensional cases will be dealt with.

4.1. 2-DIMENSIONAL CASE

First of all, it is needed to establish the geometric figure that will define the uniform distribution of points. According to [6], we have selected the circle[1].

In order to find g, let us start from the general expression of J_M

$$J_M(U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} u_{ik}^m \cdot d_{ik}^2(v_i, x_k) \tag{7}$$

considering the case of $m=1$ and X being a family of objects uniformly distributed inside the unit circle with a infinite number of elements. The first assumption implies that the solutions will be crisp clusters (see [2]) while the second assumption transforms the finite sum on k to a double integral inside the circle.

As the clusters are crisp, u_{ik} is 1 for the cluster k, 0 for the other clusters and ½ in the intersection between clusters.

[1] Another shapes (for instance, a square) will need a further study.

We have studied two situations:

1. For c between 2 and 5, the area or each cluster has the shape shown in Figure 1.
2. For c=6 and c=7, a cluster in the origin turns up. Figure 2 shows an example.

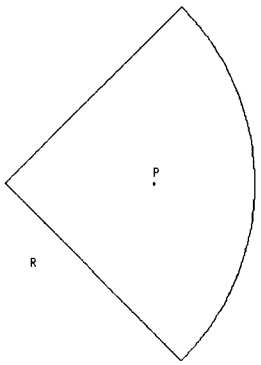


Figure 1: Shape of clusters in situation 1

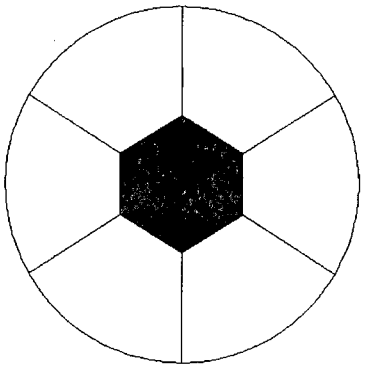


Figure 2: Regions in situation 2

Let us work in the situation 1. The formulation is (where R is the region where the cluster has points):

$$J_2(U,V) = \sum_{i=1}^{c} \iint_R d_{ik}^2(v_i, x_k) \tag{8}$$

As all the clusters are equal, we can eliminate the sum, having:

$$J_2(U,V) = c \iint_R d_{ik}^2(v_i, x_k) \tag{9}$$

Then the centroid P can be placed in the x axis, so that:

$$J_2(U,V) = c \iint_R \left((x-P)^2 + y^2\right) dxdy \tag{10}$$

Changing to polar co-ordinates:

$$J_2(U,V) = 2c \int_{r=0}^{1} \int_{\vartheta=0}^{\pi/c} (r^2 + P^2 - 2rP\cos(\vartheta)) r dr d\vartheta = \\ = \pi/2 + P^2\pi - 4cP\,\mathrm{sen}(\pi/c)/3 \tag{11}$$

The minimum of the function is reached when

$$P = \frac{2c\,\mathrm{sen}(\pi/c)}{3\pi} \tag{12}$$

In this way, we have:

$$J_2(U,V) = \pi/2 - 4c^2\,\mathrm{sen}(\pi/c)^2/9\pi \tag{13}$$

For the situation 2 the result is:

$$J_2(U,V) = \frac{\pi}{2} + P^2\pi - \frac{4(c-1)P\,\mathrm{sen}(\frac{\pi}{c-1})}{3} - \frac{(c-1)P^4\,\mathrm{tg}(\frac{\pi}{c-1})}{12} \tag{14}$$

where P can be solved in the same way as before.

To obtain the equalization function, we should solve the equation

$$g(c) = K / J_2(U,V) \tag{15}$$

Nevertheless, as the expression of J_2 is not simple, we directly use the coefficients. The next table shows the *1/g* coefficients. The need of studying situation 2 is proved due to the fact that for 6 and 7 clusters the values of (13) are greater than the values of (14).

Table 1: Coefficients for 1/g in the 2-D case

C	*2*	*3*	*4*	*5*	*6*	*7*
1/g	1.00	0.62	0.44	0.35	0.29	0.24

In the next section it will be presented some results using these coefficients.

4.2. 3-DIMENSIONAL CASE

For the 3-dimensional case, the situation is similar to the previous one studied. In this case, the reference case studied is the sphere of radius equal to 1.

The way of solving J_3 is also similar, but the situations found were different. These are the following:

1. For *c* from 2 to 4, the area of each cluster has the shape shown in Figure 3.
2. For *c* from 5 to 7, each cluster has the shape shown in Figure 4.

In the first situation the integral is the following:

$$J_3(U,V) = 2c \int_{r=0}^{1} \int_{\vartheta=\pi/2}^{\pi/2+\pi/c} \int_{\varphi=0}^{\pi} (r^2 + P^2 - 2rP\cos\vartheta\,\mathrm{sen}\,\varphi) \cdot r^2\cos\vartheta \cdot dr \cdot d\vartheta \cdot d\varphi \tag{16}$$

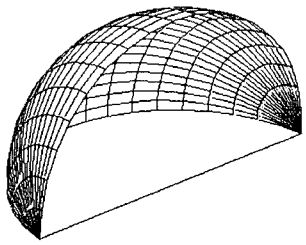


Figure 3: Shape of clusters in situation 1

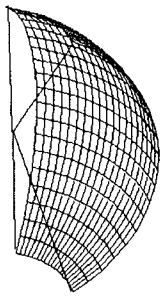


Figure 4: Shape of clusters in situation 2

It is minimized when:

$$P = \frac{3c\,\mathrm{sen}(\pi/c)}{16} \tag{17}$$

And the final expression of J_3 is:

$$J_3(U,V) = 4\pi/5 - 3\pi \cdot c^2\,\mathrm{sen}(\pi/c)^2/64 \tag{18}$$

About the situation 2, we have found that the centroids are the vertex of regular polyhedrons. These polyhedrons have symmetry with respect to a plane that crosses the center of the sphere. Then, for 6 sets the centroids are vertex of an octahedron, and the symmetry plane have four centroids that make up a square. Respectively, for 5 sets there are 3 centroids in the symmetry plane that compose an equilateral triangle and for 7 sets there are 5 centroids in the symmetry plane that make up a pentagon.

About the equalization function, g will be defined by coefficients. They are shown in the following table.

Table 2: Coefficients for 1/g in the 3-D case

c	*2*	*3*	*4*	*5*	*6*	*7*
g^{-1}	1.92	1.52	1.33	1.04	0.88	0.80

5 RESULTS

In this section it will be shown the application of the equalization function g to a practical case.

5.1. 2-DIMENSIONAL CASE

Figure 5 shows the data set used for testing. It can be seen that there are four well separate clusters. Figure 6 presents the value of J_2 normalized by g. Minimum is found in 4 clusters, as it was expected.

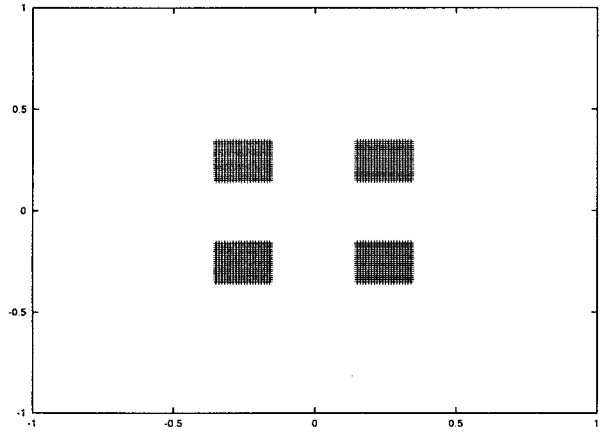


Figure 5: Data set for the 2-D case

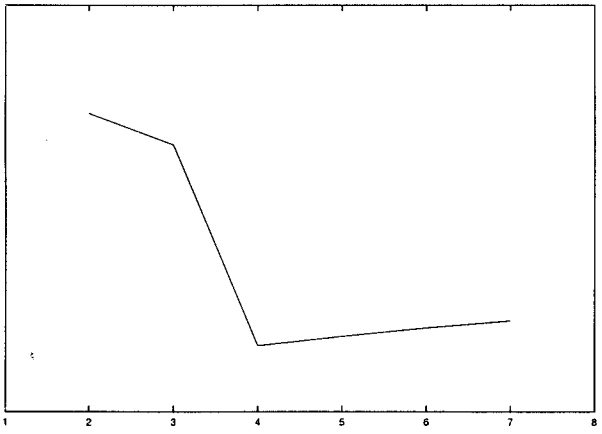


Figure 6: Results for the 2-D case

5.2. 3-DIMENSIONAL CASE

Data set used for testing is shown in Figure 7. It can be seen that there are four well separate clusters too. The value of J_3 normalized by g is presented in Figure 8. The same as before, the result of 4 clusters is correct.

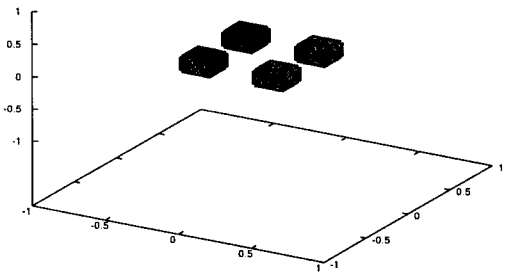


Figure 7: Data set for the 3-D case

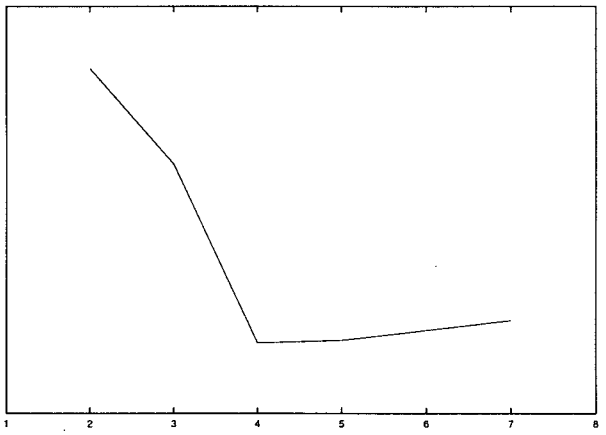


Figure 8: Results for the 3-D case

6 REFERENCES

[1] J.Dunn. A fuzzy relative of the isodata process and its use in detecting compact well separated clusters. J. Cybernetics, 3(3):32-57, 1973.

[2] S. López, L. Magdalena, J.R. Velasco. Genetic fuzzy c-means algorithm for the automatic generation of fuzzy partitions. IPMU'98.

[3] G.J. Klir, B. Yuan. Fuzzy sets and fuzzy logic theory and applications. Prentice-Hall, 1995.

[4] N.R. Pal, J. Bezdek. On cluster validity for the fuzzy c-means model. IEEE Trans. On Fuzzy Systems, vol 3, Num 3, pp. 370-379, 1995.

[5] J.R. Velasco, S. López, L. Magdalena. Genetic fuzzy clustering for the definition of the fuzzy sets. Proc. FUZZ-IEEE'97, pp. 1665-1670, 1997.

[6] M. Windham. Cluster Validity for the fuzzy c-means clustering algorithm. IEEE Trans. On Patt. Anal. And Mach. Int. vol PAMI-4, no. 4, pp. 357-363, 1982.

FUZZY CLUSTERING WITH AMBIGUITY FOR MULTI-CLASSIFIERS FUSION : CLUSTERING-CLASSIFICATION COOPERATION

Veyis Gunes
Laboratoire L3i
Université de La Rochelle
Avenue Marillac
17042 La Rochelle Cedex (FRANCE)
vgunes@univ-lr.fr

Michel Ménard
Laboratoire L3i
Université de La Rochelle
Avenue Marillac
17042 La Rochelle Cedex (FRANCE)
mmenard@univ-lr.fr

Pierre Loonis
Laboratoire L3i
Université de La Rochelle
Avenue Marillac
17042 La Rochelle Cedex (FRANCE)
ploonis@univ-lr.fr

Summary

The main aim of this paper is to demonstrate the performance of multi-classifiers fusion based on fuzzy clustering with ambiguity. The problem is seen from the multi-decision point of view (i.e. several classification modules). Each classification module is specialized on a particular region of the features space. These regions are obtained by fuzzy clustering and constitute the original data set by union.

Keywords: Fuzzy clustering, Ambiguity, Fusion, Pattern recognition.

1 INTRODUCTION

Using a pattern recognition approach, the p parameters observed build up the pattern vector. The recognition is linked to the pattern classes to be discriminated in the $p-$dimensional representation space. Real world problems are so complex that a single classifier managing whole classes and features is not well-suited. Indeed, the training data set is too large for a correct learning. For the combination of multiple classifiers, an algorithm usually take two basic approaches :

- classifier fusion algorithms. In this case, individual classifiers are applied in parallel and their outputs are combined to achieve a consensus; frameworks of these algorithms include the vote theory [10], unanimous consensus [10] [2], polling methods which utilize heuristic decision rules [4], Bayes theory [10], Dempster-Shafer theory [8] [10], possibilistic theory [1], and multistage classification [3]. These methods often provide techniques for subdividing the features space into a suitable number of sub-spaces allowing the cooperation of different classification techniques. Even if the global behavior is improved due to a kind of meta-decision frontiers design, the specification of each module on a parameter sub-space decreases each classifier recognition rate [5].

- adaptive classifier selection. Each classification module is specialized on a specific region of features space. The methods attempt to predict which single classifier is most likely to be correct for a given sample. Thus, only the output of the selected classifier is considered in the decision step. Several papers present methods to select a classifier that use estimates of each individual classifier's local accuracy in small regions of features space surrounding an unknown test sample [9]. However, only the output of the best classifier for this data is used to make the final decision. These algorithms are often based on optimization methods, as for example, genetic algorithms and neural networks [7].

It is possible to combine both approaches using methods partitionning the features space into a suitable number of regions : the goal is to allow the cooperation of several classification modules. Moreover, in order to avoid a complex conflicts managing, the fusion step must be realized with only a subset of classifiers (the most concerned by).

In this paper, a pattern recognition approach is presented. This one is based on supervised and unsupervised procedures and uses the Dempster-Shafer theory of evidence. These procedures are combined to form an adaptive classifier selection and fusion system. The system starts with a given set of training patterns labeled among the known classes. The figures 1 and 2 illustrate the method. An unsupervised mechanism generates a new partition, including ambiguity clusters. This partition is obtained using the Fc+2M algorithm [6]. A classifier is associated with each simple cluster. The training set for each classifier is then defined on the cluster and its associated ambiguous clusters (see figure 2). Each classifier is trained separately. The decision making for an unknown pattern is the fusion of outputs from *neighbour* classifiers (see section 2.2). The method can be defined as follows :

- let $\Phi = (\phi_i)_{i\in[1,l]}$ be a family of l classes ; the recognition is linked to the pattern classes to be discriminated in the p-dimensional representation space ;
- let $X = (x_k)_{k\in[1,n]}$ be the family of n objects where $x_k = \left(x_{k1}, x_{k2}, \ldots, x_{kp}\right)^t$ is a pattern described by p features (i.e. $x_k \in R^p$). Let $X^{lr} \subset X$ be a set of training patterns labeled according to the known Φ family ;
- let $\Omega = (\omega_i)_{i\in[1,c]}$ be a family of c clusters; These clusters are given by the fuzzy clustering processing applied in the p-dimensional representation space. Since there are 2^c subsets in a set of c elements, we obtain $2^c - 1$ regions, excluding the empty set, in a $c-$cluster problem. We attempt to partition the features space into regions; each of these corresponds to a subset of X^{lr}. Thus, in a $c-$cluster problem there are c simple clusters, $(\{\omega_i\})_{i\in[1,c]}$, and to exercise the ambiguity reject option, we associate an ambiguity cluster with each subset of clusters $A \in 2^{\Omega}\setminus\{\emptyset, (\{\omega_i\})_{i\in[1,c]}\}$. In figure 3 there are $2^3 - 1$ regions corresponding to the subsets $\{\omega_1\}, \{\omega_2\}, \{\omega_3\}$, $\omega_{12} = \{\omega_1, \omega_2\}$, $\omega_{13} = \{\omega_1, \omega_3\}$, $\omega_{23} = \{\omega_2, \omega_3\}$ and $\omega_{123} = \{\omega_1, \omega_2, \omega_3\}$;
- let $E = (e_k)_{k\in[1,c]}$ be a set of c classifiers computing the possibility that an unknown test sample belongs to the classes $(\phi_i)_{i\in[1,l]}$. A classifier is associated with each cluster i.e. the training data set, S_{ω_i} for the classifier e_i is defined on the cluster ω_i and its associated ambiguous clusters (see figure 2).

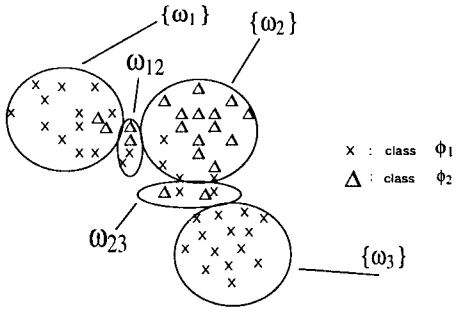


Figure 1: The partition of the sets of training patterns, $X^{lr} \subset X$, if the ambiguity reject is exercised. $\{\omega_1\}, \{\omega_2\}, \{\omega_3\}$: simple clusters; ω_{12}, ω_{23} : ambiguous clusters.

Each classification module is specialized on a specific region of the features space. These regions are obtained by fuzzy clustering and produce the original data set by union. The goal is to reduce the amount of data by merging similar patterns together into clusters. Moreover, we introduce the ambiguity notion in clustering, which deals with patterns lying near the clusters boundaries. The overall system is parallel, since different classifiers work with their own

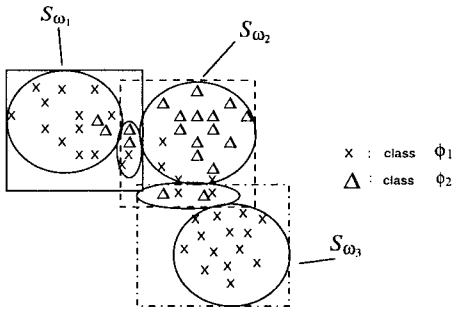


Figure 2: One classifier is associated with each cluster. The training data set, S_{ω_i} for the classifier e_i is defined on the cluster ω_i and its associated ambiguous clusters.

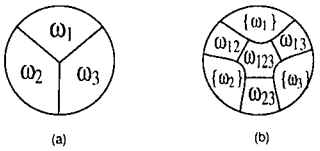


Figure 3: (a) the partition of the pattern space X into three regions corresponding to simple cluster. (b) the partition if the ambiguity reject is active. We associate an ambiguity cluster with each subset $A \in 2^{\Omega} \setminus \{\emptyset, (\{\omega_i\})_{i\in[1,3]}\}$.

training data clusters. The algorithm makes it possible the adaptive classifiers selection in the sense that the fuzzy clustering with ambiguity gives adapted training data regions of features space. The decision making is the fusion of outputs from *neighbour* classifiers.

2 A PARALLEL AND ADAPTIVE MULTI-CLASSIFIERS FUSION

The overall system consists of four major components : a set of c classifiers $(e_k)_{k\in[1,c]}$ which are responsible for calculating the possibility that a particular test pattern belongs to cluster ϕ_i, $i \in [1, l]$, a fusion module $\mathcal{F}$ which merges the outputs from subset of individual classifiers, a decision module $\mathcal{D}$ which assigns the test pattern to one of the l classes and a selection module which chooses the subset of classifiers from the set of classifiers for a given pattern.

2.1 THE CLUSTERING PROCEDURE : SELECTION MODULE

The Fc+2M algorithm generates a new partition $\mathcal{P}$ of X^{lr}, including ambiguity clusters. The number of simple clusters, c, is generally higher than the number of classes, l. The use of clustering on complex and noised data permits to specify the training sets and the neighbourhood of the classifiers, and deals with the cooperation between the classifiers allowing the multiple points of view decision man-

agement. The main idea is that the clustering procedure can manage the selection of the most adapted classifiers.

2.2 THE CLASSIFIERS

The bayesian classifier, well known statistical approach, is considered in this study. The main hypothesis is the knowledge on the distribution law whose features are determined on the learning set, X^{lr}. One classifier is associated with each simple cluster. The training data set, S_{ω_i}, for the classifier e_i is defined on the cluster ω_i and its associated ambiguous clusters : $S_{\omega_i} = \{x_k \in A \ : \forall A \in 2^{\Omega} \text{ and } \omega_i \in A\}$. The decision classes for e_i are defined by $\{\phi_j \in \Phi \ : \phi_j \ni x_k \ \forall x_k \in S_{\omega_i}\}$. Let $(e_k)_{k\in[1,c]}$ be the set of classifiers. The set of neighbouring classifiers for the classifier e_i is defined by: $E_i = \{e_k \ : S_{\omega_i} \cap S_{\omega_k} \neq \emptyset\}$.

2.3 FUSION MODULE

Information fusion is an important aspect of any intelligent system. Let $X^t \subset X (X^t \cap X^{lr} = \emptyset)$ be the set of unknown patterns. For classification, an unknown pattern $x_k \in X^t$ is assigned to the partition $\mathcal{P}$. Thus the following steps for pattern recognition at this level are :

- if x_k is classified in a simple cluster, ω_i, the output of the associated classifier, e_i, with this cluster is used to make the final decision;
- if x_k is classified in an ambiguous cluster, the decision making is expressed as an aggregation of the set of outputs from neighbour classifiers (see figure 4) such as : $\{e_j \ : \omega_j \in A\}$

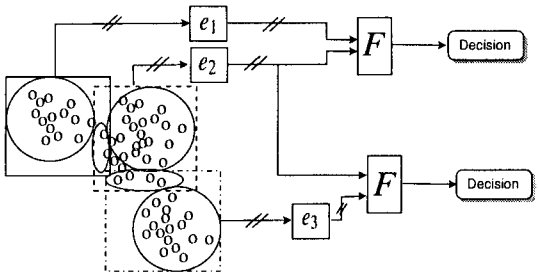


Figure 4: The decision making for an unknown pattern is expressed as an aggregation of the set of outputs from neighbour classifiers.

Several methods for combining pieces of information issued from several sources, based on certainty factors as MYCIN algorithm, or on credibility and plausibility measurements as in Evidence Theory were proposed in literature. Our proposed Fusion module is based on the combination using the Dempster-Shafer's orthogonal rule [8].

3 RESULTS

The above method, named 3C (as clustering-classification cooperation), is tested on a set of synthetic data (2-dimensional data). The learning set X is composed of 210 patterns (figure 5). Our choice is $\sigma = 2$ (standard deviation).

- Φ_1 : 100 Gaussian distributed patterns according to :

$$m_1 = \begin{pmatrix} 10 & 20 \end{pmatrix}^t, \Sigma_1 = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

- $\Phi_2 : (\Sigma_2 = \Sigma_1)$
  - Gaussian distributed, according to :

$$m_{2a} = \begin{pmatrix} 10 & 10 \end{pmatrix}^t \ 50 \, patterns$$

$$m_{2b} = \begin{pmatrix} 10 & 30 \end{pmatrix}^t \ 50 \, patterns$$

  - 10 patterns (on the right of Φ_1)

The test set is generated in the same conditions, as shown in figure 6. The Euclidian metric is used as a distance measure. Let's notice that during the clustering stage (Fc+2M), distance rejection is avoided. The results obtained are shown in figure 7 (c=2). Fc+2M parameters are : $m = 1.5$ and $\alpha = 0.1$. The partition obtained is as shown in table 1 for $c = 2$ and in table 2 for $c = 3$. All the values in the tables are expressed in percentage. The results after fusions are shown in table 3 (for $c = 2$). A comparison, with the bayesian theory of decision applied to the whole of data set, is performed. Table 4 shows the results for the three methods (bayesian, 3C algorithm with $c = 2$ and $c = 3$). The overall best performance is obtained with the 3C method for c=2.

Table 1: Fc+2M partition of X^t for $c = 2$: Φ_1 and Φ_2 are real classes; ω_1 and ω_2 are simple clusters; ω_{12} is the ambiguous cluster.

↗	ω_1	ω_2	ω_{12}
Φ_1	48.0	15.0	37.0
Φ_2	50.91	49.09	0

Table 2: Fc+2M partition of X^t for $c = 3$: There are 3 simple clusters and 4 possible ambiguous clusters (2 of them, ω_{12} and ω_{13} exist really).

↗	ω_1	ω_2	ω_3	ω_{12}	ω_{13}	ω_{23}	ω_{123}
Φ_1	67.0	0	0	17.0	16.0	0	0
Φ_2	3.6	36.4	45.5	11.8	2.7	0	0

Table 3: 3C Algorithm : Confusion matrix for $c = 2$.

$\nearrow$	Φ_1	Φ_2
Φ_1	95.0	5.0
Φ_2	0.91	99.09

Table 4: Comparison with bayesian classifier.

	Bayes	3C (c=2)	3C (c=3)
Success	88.6	97.1	96.7
Confusion	11.4	2.9	3.3

4 CONCLUSION

In this paper, we have shown that the ambiguity reject in the clustering stage is very useful in order to constitute subsets of data for the learning stage of more localized set of bayesian classifiers. Our results show that in the case of non-gaussian classes, a much better classifier than a single global bayesian classifier can be obtained by using multiple local bayesian classifiers associated with fusion by Dempster-Shafer orthogonal rule. In this latter stage, ignorance could be calculated by using rates of patterns affected to each single classifier.

References

[1] D. Dubois and H. Prade. Possibility theory. *Plenum Press. New York*, 1988.

[2] T. K. Ho, J.J. Hull, and S.N. Srihari. Decision combination in multiple classifier systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16:66–75, jan 1994.

[3] Y.S. Huang and Suen C.Y. A method of combining multiple classifiers - a neural network approach. *Proc. 12th Int'l Conf. Pattern Recognition and Computer Vision. Jérusalem*, pages 473–475, 1994.

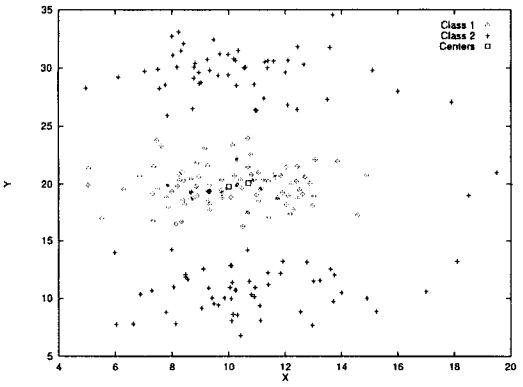


Figure 5: *Learning set:* $X^{lr} \subset X$

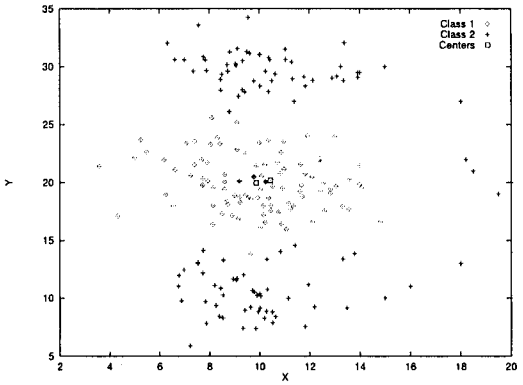


Figure 6: *Test set:* $X^t \subset X$

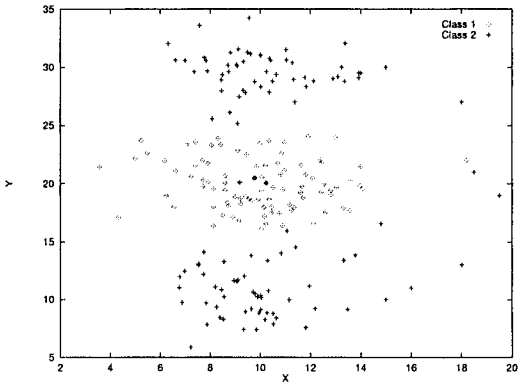


Figure 7: *Results obtained with 3C algorithm*

[4] F. Kimura and M. Shridar. Handwritten numerical recognition based on multiple algorithms. *Pattern recognition*, 24(10):969–983, 1991.

[5] P. Loonis and M. Ménard. z_{01}, a reliability limit in pattern recognition: Application in exterior scenes area and automatic fish sorting. In *Proceedings of IPMU*, Paris, july 1998.

[6] M.Menard, C.Demko, and P.Loonis. The fuzzy c+2-means: Solving the ambiguity reject in clustering. In *Pattern recognition. To be published.*

[7] P.Loonis, M.Ménard, and C.Demko. A new genetic algorithm for the multi-classifiers fusion optimization. In *Proceedings of IPMU*, Granada, Spain, July 1996.

[8] G. Shafer. A mathematical theory of evidence. *Princeton University Press,Princeton, N.J.*, 1976.

[9] K. Woods, W.P. Kegelmeyer, and JR.K. Bowyer. Combination of multiple classifiers using local accuracy estimates. In *PAMI*, April 1997.

[10] L. Xu, A. Krzyzak, and C.Y. Suen. Methods of combining multiple classifiers and their applications to handwritting recognition. *IEEE Trans. Systems Man and Cybernetics*, 22(3):418–435, 1992.

FUZZY LOCAL REGRESSION MODELS WITH FUZZY CLUSTERING *

A. F. Gómez-Skarmeta, H. Martínez, J. Botía Blaya
Dept. Informática, I.A. y Electrónica,
Facultad de Informática, Univ. de Murcia,
e-mail:skarmeta@dif.um.es

M. Delgado
Dept. Ciencias de la Computación e I.A.,
E.T.S.I. Informática, Univ. de Granada,
mdelgado@ugr.es

Summary

In this work we present an alternative approach to generate the fuzzy rules with multidimensional fuzzy sets in the antecedent and functional consequent associated to the TSK fuzzy model. In order to obtain the fuzzy rules we use a fuzzy clustering algorithm that look for linear behaviours in the product space of the input-output data.

Keywords: fuzzy clustering; unsupervised learning; fuzzy modeling.

1 INTRODUCTION

The TSK model introduced by Takagi, Sugeno and Kang (TSK fuzzy reasoning) [10] is associated with fuzzy rules that have a special format with a functional type consequent instead of the fuzzy consequent that normally appears in the Mamdami Model. In this way the TSK approach tries to decompose the input space into subspaces and then approximate the system in each subspace by a simple linear regression model. This characteristic provides efficient models to deal with complex system, although the generation of the corresponding fuzzy rules, specially the premise structure, is technically difficult and may lead to a nonlinear programming problem. Several alternative approaches have been proposed to reduce the complexity of this building process.

In the area of fuzzy modeling the procedure of fuzzy clustering has been utilized in different ways. In one of them fuzzy clusters give rise to "local" regression models - this is in fact the essence of the idea introduced originally in [10] and that have been used in differents works like [9][12]. The overall model is then structured into a series of IF-THEN statements. The condition part of the statement might includes linguistic labels of the input variables, or directly fuzzy relations on the product space of the input variables, while the action part contains a linear (or generally nonlinear) numerical relationship between input and output variables.

In this work we present an alternative approach to generate the fuzzy rules with functional consequent associated to the TSK fuzzy model. In our case using fuzzy clustering algorithms that look for linear behaviours in the product space of the input-output data, we develop an approach that detect multidimensional fuzzy sets in the product space of the input variables and in this way trying to identificate the premise of the fuzzy rules and then assigning a linear consequent to each rule.

The identification of fuzzy models can be improved in this way using these multidimensional reference fuzzy sets . Other authors work with fuzzy clusters in the product space of the input-output space using in this way the clustering in order to detect the interaction between the input and output variables, hence fuzzy clusters give rise to "local" regression models . The model is then structured into a set of IF-THEN statements. The condition part of each rule is obtained by projecting the fuzzy clusters in the input and output spaces as in [5]. Several of this authors have proposed the use of the fuzzy clustering techniques that tend to find clusters that are more adequate to the geometry of fuzzy rules than the classical fuzzy c-means [3] [7]. The most significant difference with other works is our use of a Gustafson-Kessel clustering algorithm in the product space of the input-output space in order to identify the fuzzy rules and then use it to generate the fuzzy rules with functional consequent within the TSK fuzzy model.

The paper is organized as follows. In section 2 we formulate the fuzzy model with which we are going to

*This work has been partially supported by CICYT project TIC97-1343-C002-02 and FEDER and EU project 1FD97-0255-C03-01

work. In section 3 we define the methods that we propose here to use the fuzzy clusters in order to generate the fuzzy rules of the TSK model. Numerical examples are provided in section 4 to illustrate the performance of the presented methods. Finally, section 5 contains some conclusions and comments about future trends.

2 FUZZY MODELS AND FUZZY CLUSTERING

Usually the basic knowledge to model a (fuzzy) system is assumed to be given in terms of a (usually large) set of sample I/O pairs, say $\Omega = \{(x_{t1}, x_{t2}, \ldots, x_{tp}), y_t)\} = \{(x_t, y_t), t = 1, \ldots, N\}$, and obviously the selected rules are required to approximate the function $\varphi : X^p \rightarrow Y$ that theoretically underlies the system behaviour.

We assume that the collection of data samples represent the system behavior in the product space $(X^p \times Y)$, where $X^p = (X_1 \times X_2 \times \ldots \times X_p)$, with $X_1, X_2, ..., X_p$ being the domains of discourse of the inputs and Y is the domain of the output.

It is clear that the approximation of the function $\varphi : X^p \rightarrow Y$ it is equivalent to the establishment of a relation between the data in X^p and Y, which in the fuzzy modeling context will be a fuzzy relation in the product space of the input-output data In this way, the clustering of the data in $(X^p \times Y)$ seems to be an adequate technique to disclose this relation and several authors, like Yoshinari et al.[12] Yager and Filev[11], Babuska and Verbruggen[1] among others, have use it in different ways to identify fuzzy rules. Our contribution is in the use of the information obtained from the clustering, in a direct way in order to generate the fuzzy modeling.

Let us remark that in the TSK model we need to specify fuzzy sets only in the input universe of discourse, so that the antecedent of each rule define a fuzzy region in the product space of the input variables X^p. Several authors (Sugeno et al.[9], Yager et al.[11], Babuska et al.[1] etc.) have made use of fuzzy clustering in fuzzy modeling by projecting the clusters in the domains of the variables. We have adopted a different point of view: instead of specifying fuzzy sets in each universe of discourse, we directly estimate the fuzzy region of the input space by using the fuzzy relation in the input-output product space obtained from the fuzzy clustering. In general, we will directly characterize fuzzy sets in X^p, thus trying to capture most of the information about the system behaviour. The centroids of the clusters (which describe a local behaviour) and their relation with the data (the fuzzy sets they generate), play a key role in our process.

Hence, what we need is a mean to establish a relation between fuzzy clusters in $(X^p \times Y)$ and associated fuzzy sets in X^p, as will be explained in the next section.

Let suppose that by using the fuzzy clustering we have already disclosed k local behaviour of the data (see [3] in order to detect the optimal number of fuzzy clusters) around the obtained centroids as captured by the fuzzy relations. By using the information the clusters in $(X^p \times Y)$ represent (centroids and distribution of the data around them), we are going to assign to any data x of X^p a value in Y, so we approximate the function $\varphi : X^p \rightarrow Y$, being able to generate k fuzzy rules associated with the fuzzy clusters:

$$\begin{array}{l} R_h : \text{ If } x \text{ is } A_h \text{ then } y \text{ is } f_h(x) \\ \text{where} \\ f_h(x) = a_{0h} + a_{1h}x_1 + ... + a_{ph}x_p \\ \text{with } h = 1, ..., k \end{array} \qquad (1)$$

where $x = (x_1, ..., x_p)$ is the whole input variable and A_h fuzzy sets in X^p to be obtained from the hth fuzzy cluster C_h of $(X^p \times Y)$.

3 FUZZY LOCAL REGRESSION MODELS

Instead of using directly a classical fuzzy clustering algorithm like the Fuzzy C-Means [2], as is usual in the literature - an algorithm that is one of the most popular - we have adopted an alternative approach. We have used the Gustafson-Kessel fuzzy clustering algorithm [4] starting from the results of a previous application of a FCM method. This algorithm is based on the detection of hyperplane fuzzy clustering and thus it is adequate to detect the fuzzy partitions that better fulfill the assumption of fuzzy linear models [1]. The objective function of this algorithm is similar to the FCM, although it admits different shape and size in the fuzzy clusters.

$$J(U, V : X) = \sum_{i=1}^{N} \sum_{h=1}^{k} \mu_{hi}^m D_{hi,A_h}^2 \qquad (2)$$

where N is the number of samples, k the number of fuzzy clusters that are established using a cluster validity measure, m is a smoothing parameter which controls the "fuzziness" of the clusters (normally between $1.75 - 2.5$), and $J(\cdot)$ the objective function that must be minimized with respect to the fuzzy partition U the centroids V and the data set X, and whose solution will be determined by solving by means of a Picard iteration the expressions:

$$v_h = \frac{\sum_{i=1}^{N} \mu_{hi}^m x_i}{\sum_{i=1}^{N} \mu_{hi}^m} \text{ with } h = 1, ..., k$$

$$\mu_{hi} = \frac{1}{\left[\sum_{j=1}^{k} \left(\frac{D_{hi,A_h}}{D_{ji,A_j}}\right)^{\frac{2}{m-1}}\right]} \text{ with } i = 1, ..., N \; h = 1, ..., k$$

$$D^2_{hi,A_h} = (X_i - V_h)^T M_h (X_i - V_h)$$

$$M_h = (\rho_h \det(C_{fh}))^{\frac{1}{p}} C_{fh}^{-1}$$

$$C_{fh} = \frac{\sum_{i=1}^{N} \mu_{hi}^m (X_i - V_h)(X_i - V_h)^T}{\sum_{i=1}^{N} \mu_{hi}^m}$$

where v_h are the centroids of the fuzzy clusters, μ_{hi} is the membership degree of (x_i, y_i) to the $h-$th cluster C_h, C_{fh} is the fuzzy covariance matrix of this C_h. The optimal value would be such that $\det\left(M^*_{fh}\right) = \rho_h$

Once we have found the fuzzy clusters in $(X^p \times Y)$ by the Gustafson-Kessel algorithm, we need to determine A_h and $f_h(\cdot)$ of expression 1.

The fuzzy clusters in the product space $(X^p \times Y)$ give information about how the data are structured in X^p captured in the centroids and the metric found by the clustering on the product space, which is to be used to induce fuzzy regions in X^p. We define fuzzy sets with exponential membership functions in the domain X^p, using the components in this space of the centroids found in $(X^p \times Y)$ and the fuzzy covariances of the sample data components in X^p. The centroids of the clusters found in $(X^p \times Y)$ are noted as c^h_{XY}, and the components of these centroids in X^p as c^h_X.

Once we have found the fuzzy clusters in $(X^p \times Y)$ using the Gustafson-Kessel algorithm, one important difference of our approach from the classical TSK model, is that we will use directly the grade of membership of the data to the fuzzy clusters found in $(X^p \times Y)$, in order to obtain by means of the recursive least-squares algorithm (or a stable-state Kalman Filter), the coefficients of the linear regression model that will be associated to the fuzzy region corresponding to each fuzzy cluster.

This way we obtain a method:

$$GMBD \equiv \hat{y} = \frac{\sum_{h=1}^{k} \exp\left(-\frac{\|x - c^h_X\|^2_{C_{fh}}}{2}\right) * f_h(x)}{\sum_{h=1}^{k} \exp\left(-\frac{\|x - c^h_X\|^2_{C_{fh}}}{2}\right)} \quad (3)$$

where C_{fh} is the covariance matrix over the components in X^p associated with the hth cluster, and $f_h(\cdot)$ is the linear function with coefficient obtained by means of the recursive least-squares algorithm using the grade of membership of the data to the fuzzy clusters found in $(X^p \times Y)$.

The motivations for using exponential membership functions are that these membership functions are symmetric under rotation in the multidimensional space and therefore can be easily projected onto each domain of the variables. These projections can be used to generate fuzzy rules with fuzzy sets in each domain of discourse, this kind of fuzzy rules being more appropiate for linguistic interpretations.

4 NUMERICAL EXAMPLES

In this section we present the result of the different methods applied to two different examples: A) the Box-Jenkins gas furnace and B) the Zimmermann and Zysno model based on empirical data [13].

The performance index considered to evaluate the obtained fuzzy models will be the root mean square error:

$$E = \sqrt{\frac{\sum_{t=1}^{N} (y_t - \hat{y}_t)^2}{N}} \quad (4)$$

where N is the number of data, y_t is the actual output, and $\hat{y}_t$ is the model output.

4.1 BOX-JENKINS GAS FURNANCE

Box and Jenkins's gas Furnace is a famous example of system identification. The well-known Box-Jenkins data set consists of 296 input-output observations, where the input $u(t)$ is the rate of gas flow into a furnace and the output $y(t)$ is the CO_2 concentration in the outlet gases. Since the process is dynamical, there are differents values of the variables that can be consider as candidates to affect the present output $y(t)$. In our case we have considered as inputs $y(t-1)$ and $u(t-4)$ as different authors have done [12], etc.

In table(1) we present the comparison of our method (GMBD) with several fuzzy models proposed in the literature, showing the number of variables chosen in each case, the number of rules and the performance of the models. In order to compare our results we carry out a fuzzy clustering on the 296 data points, considering 2 clusters.

We obtain a better performance with a number of rules remarkably smaller in comparison with other "classical" approaches. The Sugeno and Tanaka model, Sugeno (91) provide smaller error, although with more variables and more computational complexity.

4.2 ZIMMERMANN AND ZYSNO EMPIRICAL DATA

In this section we present the results of our method on the empirical data presented in [13] and that other

Table 1: Errors in Literature

Model	N. Inputs	N. Rules	Error
Tong(77)	2	19	0.684
Pedrycz(84)	2	81	0.565
Xu(87)	2	25	0.572
Peng(88)	2	49	0.548
Sugeno(91)	6	2	0.261
Sugeno(93)	3	6	0.435
Wang(96)	2	5	0.397
GMBD	2	2	0.396

authors like Dyckhoff and Pedrycz, Krishnapuram and Lee [6], or Langari and Wang [8] have also used to test their algorithms.

Table(2) shows the performance of the method suggested in the literature and our methods (with four rules as suggested by Langari et al.[8].)

Table 2: Errors in Literature

Model	PI/Error
Zimmermann and Zysno(80)	0.057
Dyckhoff and Pedrycz(84)	0.052
Krishnapuram and Lee(92)	0.051
Langari and Wang(96)	0.042
GMBD	0.034

Again in this example it is clear the good performance of our method.

5 CONCLUSIONS AND FUTURE TRENDS

The use of a fuzzy clustering algorithm like fuzzy C-means(FCM) in combination with the Gustafson-Kessel algorithm, results into an efficient method to identify a collection of fuzzy rules. That suggest us the future analysis of other kind of combinations between fuzzy clustering algorithms. Like all unsupervised techniques, clustering suffers from the presence of noise in the data. While this is not really a drawback when nothing is known about the nature and number of subgroups in the data set, it can be a serious problem in situations where one whishes to generate membership functions from training data. In order to treat this problem, the use of fuzzy clustering algorithm like the Possibilistic C Means (PCM) or its variants could be an interesting alternative.

References

[1] R. Babuska and H. B. Verbruggen, Applied fuzzy modeling, *Proceedings IFAC congress AIRTC,* Valencia, Spain, 1994.

[2] J. C. Bezdek, Pattern recognition with fuzzy objective function algorithms, Plenun, New York, 1981.

[3] M. Delgado, A. F. Gómez-Skarmeta and F. Martín Using fuzzy clusters to model fuzzy systems in a descriptive approach , *Proceedings of the Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU'96,* , Granada, Spain, 1996.

[4] D. E. Gustafson and W. C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, *Proc. of the IEEE Int. Conf. on Fuzzy Systems,* San Diego, pp. 761-766, 1979.

[5] F. Klawonn and R. Kruse, Constructing a fuzzy controller from data, *Fuzzy Sets and Systems,* vol. 85, pp. 177-193, 1997.

[6] R. Krishnapuram and J. Lee, Fuzzy connective-based hierarchical aggregation networks for decision making, *Fuzzy Sets and Systems*, vol 46, pp. 11-27, 1992

[7] A. Kroll, Identification of functional fuzzy models using multidimensional reference fuzzy sets, *Fuzzy Sets and Systems*, vol. 80, pp. 149-158, 1996

[8] R. Langari and L. Wang, Fuzzy models, modular networks, and hybrid learning, *Fuzzy Sets Syst.*, vol. 79, pp. 141-150, 1996

[9] M. Sugeno and T. Yasukawa, A Fuzzy-logic-based approach to qualitative modeling, *IEEE Trans. Fuzzy Systems*, vol 1, n.1, pp. 7-31, 1993.

[10] T. Takagi and M Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Transactions on Systems, Man, and Cybernetics,* vol. SMC-15, n. 1, pp. 116-132, 1985.

[11] R. R. Yager and D. P. Filev, Unified structure and parameter identification of fuzzy models, *IEEE Trans. on System Man and Cybernetics*, vol. 23, pp. 1198-1205, 1993

[12] Y. Yoshinari, W. Pedrycz and K. Hirota, Construction of fuzzy models through clustering techniques, *Fuzzy Sets and Systems,* 54, pp. 157-165, 1993.

[13] H-J. Zimmermann and P. Zysno, Latent connective in human decision making, *Fuzzy Sets and Systems*, vol 4, pp. 37-51, 1980

AUTHOR INDEX

A

Aedo, I. 291
Akdag, H. 175
Al-Hadithi, B.M. 135
Albertos, P. 99
Alcalá, R. 103
Alsina, C. 425
Aniorte, A. 287
Aracil, J. 111
Aranguren, G. 51
Arenas, M. 179

B

Barro, S. 167, 171
Benítez, J.M. 183, 473
Benvenuti, P. 367
Bernad, P. 307
Bertoluzza, C. 231
Bilbao, A. 179
Blanco, I. 493
Bodenhofer, U. 215
Bordogna, G. 489
Borgi, A. 175
Bosc, P. 243
Botía, J. 509
Bouchon-Meunier, B. 127, 139
Bugarín, A. 167, 171
Burillo, P. 395, 433, 445
Burusco, A. 247

C

Cadenas, J.M. 271
Calvo, O. 95
Calvo, T. 363, 379
Canet, P. 405
Cañada, J. 83
Cariñena, P. 167, 171
Carmona, P. 79
Casasnovas, J. 405
Casillas, J. 103
Casparri, M.T. 315
Castiñeira, E. 239, 417
Castro, J.L. 473
Chamorro, F. 449
Conejo, R. 409
Cordón, O. 159, 163, 335
Couso, I. 227
Cubero, J.C. 485, 493
Cubillo, S. 417
Cuenca, F. 493
Cuesta, F. 111

D

De Baets, B. 47, 351, 387
De Caluwe, R. 481
De Meyer, H. 387
De Moya, F. 335
De Pedro, T. 123
De Tré, G. 481
Delechamp, J. 139
Delgado, M. 279, 391, 509
Di Nola, A. 33
Díaz, I. 275
Díaz, P. 291
Divari, M. 367
Doulgeri, Z. 71
Duarte, O.G. 251, 391
Dvorák, A. 199

E

Echanobe, J. 255
Elorza, J. 395
Esteva, F. 35, 47, 299, 421

F

Faratin, P. 155
Fathi, M. 263
Fedrizzi, M. 17
Fedrizzi, M. 17
Felix, R. 25, 461

Fernández, C. 291
Fernández, F. 191
Fernández, J.M. 183
Ferreiro, R. 115, 119
Flores-Sintas, A. 271
Flórez, R. 323
Fodor, J. 47, 359
Frago, N. 445
Frías, E. 191
Fuentes, R. 247, 445

G

Galindo, J. 485
García, E. 51
García, F. 449
García, P. 421
García, P.S. 315
García, R. 123
García Carbajal, S. 331
García Galán, S. 195
García-Cerezo, A. 63
Garitagoitia, J.R. 255
Garmendia, L. 235
Garriga-Berga, C. 143
Gil, M.A. 403
Gil, P. 219, 223, 227
Gisin, V. 413
Godo, L. 35, 47, 343, 347, 421
Gómez-Skarmeta, A.F. 91, 131,509
González, A. 103
González, J.R. 255
Gordillo, F. 111
Gottwald, S. 303
Grigolia, R. 33
Gunes, V. 505
Gutiérrez, J. 191

H

Heider, R. 469
Heredia, G. 477
Herencia, J.A. 437
Hernández, E. 203
Hernández, L. 283
Herrera, F. 159, 163, 453
Hildebrand, L. 263

I

Iglesias, R. 167

J

Jacas, J. 211
Jedelsky, D. 199
Jennings, N.R. 155
Jiménez, A. 135
Jiménez, J. 231
Jiménez, M. 179

K

Kacprzyk, J. 21, 457
Keller, A. 497
Klawonn, F. 497
Klement, E.P. 371
Komorníková, M. 355
Kuijken, L. 429

L

Lamata, M.T. 437
Larsen, H.L. 327
Lazzari, L.L. 315
Linares, L.J. 441
Loonis, P. 505
López, E. 323
López, E.M. 287
López, S. 501
López-Baldán, M.J. 63
López-Díaz, M. 403
López de Màntaras, R. 295
López-Sánchez, M. 295
Lozano, J.L. 107

M

Madroño, D. 307
Magdalena, L. 83, 501
Mandow, A. 63
Mantas, C.J. 473
Marcek, D. 187
Mares, M. 375
Marichal, J.L. 383
Marín, L. 107, 399
Marques, R.A. 17
Marsala, C. 139
Martín, F. 271
Martín, J. 379
Martín-Bautista, M.J. 327
Martínez, H. 91, 509
Martínez, L. 453
Mas, M. 339
Matellán, V. 291
Matía, F. 135
Matos, N. 151
Mayor, G. 339, 379
Medina, J.M. 485
Mélin, C. 87
Mellouli, N. 127
Ménard, M. 505
Mendaña, C. 323
Mesa, F. 107, 399
Mesiar, R. 139, 351, 363, 371
Miranda, P. 219, 223
Molina, J.M. 259, 275, 449
Montes, S. 227, 231
Moraga, C. 469
Morales, R. 409
Moser, B. 75
Mucientes, M. 167, 171

N

Naessens, H. 387
Navara, M. 75
Nettleton, D. 283
Niittymäki, J. 67
Nozal, L. 51
Novák, V. 43

O

Olivares, M. 99
Olivas, J.A. 441
Ollero, A. 111, 477
Ortiz, C. 311

P

Pap, E. 371
Pardo, X.C. 115, 119
Pasi, G. 489
Peregrín, A. 163
Pérez, G. 251
Pérez, J.L. 409
Perfilieva, I. 39
Picos, R. 95
Pivert, O. 243
Pons, O. 485, 493
Portilla, M. 433
Pradera, A. 239
Puyol-Gruart, J. 299

R

Recasens, J. 203, 207, 211
Regueiro, C.V. 167
Requena, I. 183, 391
Reyes, J.A. 299
Rifqi, M. 139
Rodríguez, G. 95
Rodríguez, M. 51
Rodríguez, M.A. 323
Rodríguez, M.V. 179
Rodríguez, R.O. 421
Rodríguez-Castaño, A. 477
Rodríguez-Muñiz, L.J. 403
Rodríguez de Soto, A. 207
Roubens, M. 383
Ruan, D. 59
Ruiz, R. 287

S

Saade, J.J. 55
Sala, A. 99
Salvador, A. 235
Salvador, L. 307
Sánchez, L. 331
Sánchez, M. 91
Sánchez Álvarez, D. 131
Sánchez Fernández, D. 279
Sanchis, A. 259, 449
Sancho-Royo, A. 29
Shapiro, D. 465
Sierra, C. 151, 155, 295
Soria, R. 259
Suñer, J. 339
Szmidt, E. 457

T

Temme, K. H. 469
Theocharis, J.B. 71
Torra, V. 147, 347
Torrens, J. 379
Torres, E. 219, 223
Triguero, F. 409
Trillas, E. 235, 239, 417, 425
Turunen, E. 67

V

Valls, A. 147
Velasco, J.R. 195, 501
Velasco, M. 275
Verdegay, J.L. 29, 267
Vergara, E.R. 267
Vidal, J. 115, 119
Vidolov, B. 87
Vila, M.A. 279, 327
Villar, P. 159
Vivona, D. 367
Vojtás, P. 319

Z

Zadrozny, S. 21
Zapico, A. 343
Zarco, C. 335
Zurita, J.M. 79