Mathware & Soft Computing 8 (2001) 203-216

Relevance and Redundancy in Fuzzy Classification Systems

A. del Amo[1], D. Gómez[1], J. Montero[1] and G. Biging[2]
[1]Faculty of Mathematics. Complutense University. Madrid, Spain
e-mail: monty@mat.ucm.es
[2] Department of Environmental Sciences. University of California
Berkeley, California. *e-mail: biging@nature.berkeley.edu*

Abstract

Fuzzy classification systems is defined in this paper as an aggregative model, in such a way that Ruspini classical definition of fuzzy partition appears as a particular case. Once a basic *recursive* model has been accepted, we then propose to analyze relevance and redundancy in order to allow the possibility of *learning* from previous experiences. All these concepts are applied to a real picture, showing that our approach allows to check quality of such a classification system.

Key words: Ruspini partition, recursive system, relevance, redundancy.

1 Introduction.

From the very beginning of Fuzzy Sets Theory (see Zadeh [32]), fuzzy uncertainty have been widely considered in classification and control problems (see, e.g., Bezdek [8] and Zadeh [33]). In fact, quite often fuzziness is associated to complexity, in the sense that a fuzzy approach allows a useful simplification of the amount of information to be managed by decision makers. But sometimes concepts taken into account by decision makers are fuzzy in nature. This is the case when concepts are subject to a natural degree of membership, a standard situation in most human classification problems. Considering crisp classes instead of the true fuzzy class may appear then as a tremendous oversimplification of reality, probably misleading our efforts towards a right understanding of the real problem we are being faced to.

In many classification models it is already assumed the existence of a family of classes, previously defined. But the most interesting problems use to appear when those classes are still to be defined. From a statistical perspective, several classical *clustering* techniques have been developed. Anyway, getting a clear idea of what a class is use to require a long learning process, starting from a first proposal which is again and again modified. This argument applies both to crisp and fuzzy classification.

In fact, and denoting X the family of objects to be classified, fuzzy classification models quite often assume a previously defined countable family of fuzzy classes $\mathcal{C}$, in such a way that the key objective is to estimate the degree of membership $\mu_c(x)$ of each object $x \in X$ for each class $c \in \mathcal{C}$. In this way we are defining a membership function

$$\mu_c : X \to [0,1]$$

for each class $c \in \mathcal{C}$ (see, e.g., Roubens [21]).

Some times it is claimed that each membership function μ_c represents an isolated estimation problem, so its evaluation can be solved without taking into account other classes in $\mathcal{C}$. From our point of view, such an approach is quite unrealistic, since most decision makers use to assign degrees of membership having a previous look to all possible classes: we can not accept independence of alternatives within classification. Classification is highly dependent on the family of classes being actually shown to decision makers. Even in a crisp framework, decision makers like to have a look to all potential alternatives before assigning a particular class. It is essential to develop classification procedures based on the whole set $\mathcal{C}$ of classes under consideration, and most probably, taking into account some basic structure between classes ("big", "medium" and "small" are more than just three classes, but a structured family of classes: there is an obvious linear order explaining relationship between them).

A standard structure in practice, perhaps the simplest one, in related to the idea of *partition*. A key issue in classification modeling is indeed the notion of *partition*: a minimal structure for our family of classes under consideration. In a standard crisp classification model, for example, it is assumed that each object belongs to one and only one class (but it is not usually assumed that a given class is *in between* two other classes).

The definition of *fuzzy partition* was introduced by Ruspini [23] (see also Bezdek-Harris [7]). In his definition, it was assumed that

$$\sum_{c \in \mathcal{C}} \mu_c(x) = 1$$

for every object $x \in X$. Each one of these objects belongs to a certain degree to each class, and the total degree of membership is distributed among all classes, with no superfluous membership. The notion of crisp partition is being obviously generalized, since

$$\forall x \in X \Rightarrow \exists c \in \mathcal{C} \quad / \quad \mu_c(x) = 1, \quad \mu_k(x) = 0, \quad \forall k \neq c$$

whenever $\mu_c(x) \in \{0,1\}$, $\forall x, \forall c \in \mathcal{C}$.

Ruspini's definition may be understood as a desirable fuzzy classification system, but it should be considered too restrictive for most real implementations. It is a fact that most decision makers are not able to define such a fuzzy partition, at least at a first stage, and such a difficulty can not be simply addressed to error measurements. Most of the time the family of classes under consideration are far away from defining such a partition. It is only after a long learning process that decision maker will be able to define such a nice family of classes, holding Ruspini's condition. It is usually only after a long experience that we can define a family of classes in such a way that every object is being fully explained with no superfluous information.

Of course, it may be the case that such a Ruspini's fuzzy partition can not be reached by decision makers (such a nice family of classes may not exist or may not come to decision maker's mind). In addition, it should be pointed out that it is not true that every decision maker is willing to work with Ruspini's fuzzy partitions. For example, some fruit classification procedures are designed allowing strong overlapping between classes: due to market restrictions, it is required that each piece of fruit can be simultaneously assigned to several classes (with degrees of membership close to 1 for several classes). As pointed out in Medasani *et al.* [19], there is no single best method. Most complex problems require several simultaneous alternative approaches in order to capture and understand the limits of a chosen classification method.

Some of the above difficulties that may appear in the real implementation of Ruspiní's definition can be partially solved by means of weaker approaches (see, e.g., Thiele [27, 28]). Our proposal below (see [3]) suggests to explain fuzzy classification system by means of an aggregation model. Ruspini's definition will appear as a particular case that can be pursued while still being able to classify. Ruspini's partition is sometimes desirable but it is not a must in practice.

2 Recursive aggregation models.

Classification usually assumes a previous aggregation procedure, in order to amalgamate available information about each item relative to every class under consideration.

In a fuzzy framework, aggregation can be modeled according several kinds of fuzzy connectives: OWA operators [29], t-norms and t-conorms [25] or uninorms [6] (see also [10, 11, 12, 13, 20, 30, 31]). A criticism to some of these fuzzy connectives was addressed in Cutello-Montero [9] (see also [4]): most of those operators are either binary and associative or the number of aggregated items has to be fixed in advance. As pointed in [9], associativity is a need whenever we are willing to evaluate an arbitrary number of item based upon a unique binary operator. But it is easy to realize that binary operators evolve in time, and the overall aggregation rule still keeps consistency. *Recursiveness*, as introduced in [9], allows to avoid such a criticism, developing a formal definition of consistency for aggregation rules (see also [17] and [18] for interesting alternative approaches).

First of all, a *rule* ϕ is understood in [9] as a *consistent* family of connectives

$$\{\phi_n : [0,1]^n \to [0,1]\}_{n>1}$$

being each operator ϕ_n the one to be applied when the number of items to be aggregated is n. By definition it is assumed that connectives are continuous and non-decreasing in each coordinate (see also [17]). Such a rule will be considered *operational* in case there is an alternative calculus based upon binary operations, needing perhaps a previous re-arrangement of data.

Definition 1 *Let us denote*

$$\pi_n(a_1, a_2, \ldots, a_n) = (a_{\pi_n(1)}, a_{\pi_n(2)}, \ldots, a_{\pi_n(n)})$$

An ordering rule π *is a* consistent *family of permutations* $\{\pi_n\}_{n>1}$ *such that for any possible finite collections of numbers, each extra item* a_{n+1} *is allocated keeping relative positions of items, i.e.,*

$$\pi_{n+1}(a_1, a_2, \ldots, a_n, a_{n+1}) = (a_{\pi_n(1)}, \ldots, a_{\pi_n(j-1)}, a_{\pi_{n+1}(j)}, a_{\pi_n(j+1)} \ldots, a_{\pi_n(n)})$$

for some $j \in \{1, \ldots, n+1\}$.

Of course, data can be kept in its *natural* sequence, just in the order each one is reaching to us, real time: this is the *identity ordering rule.*

Definition 2 *A left-recursive connective rule is a family of connective operators*

$$\{\phi_n : [0,1]^n \to [0,1]\}_{n>1}$$

such that there exists a sequence of binary operators

$$\{L_n : [0,1]^2 \to [0,1]\}_{n>1}$$

verifying

$$\begin{aligned} \phi_2(a_1, a_2) &= L_2(a_{\pi(1)}, a_{\pi(2)}) \\ \phi_n(a_1, \ldots, a_n) &= L_n(\phi_{n-1}(a_{\pi(1)}, \ldots, a_{\pi(n-1)}), a_{\pi(n)}) \quad \forall n > 2 \end{aligned}$$

for some ordering rule π.

Right recursiveness can be analogously defined, and then we talk about a *recursive rule* when both left and right representations hold for the same ordering rule. *Commutative* recursive rules will be those recursive rules leading always to the same result for any possible ordering rule. We talk about *standard* recursive rules when they are based upon the identity ordering rule.

In this way recursiveness generalizes the property of associativity, still assuring an operational constructive consistency, but allowing to explain key rules in practice, like the *mean rule.* Key results concerning representation of fuzzy recursive rules can be found in [4]. They refer to a the additive solution of the general associativity equation of [16] (see also [15]).

3 Recursive classification systems.

A *recursive classification system* is a countable family $\mathcal{C}$ of fuzzy classes, in such a way that each $c \in \mathcal{C}$ has a membership function

$$\mu_c : X \to [0,1]$$

together with a *recursive triplet*

$$(\{\phi_n\}_{n\geq 2}, \{\varphi_n\}_{n\geq 2}, N)$$

where

- $\phi = \{\phi_n\}_{n\geq 2}$ is a standard recursive rule playing a *disjunction* role,
- $N : [0,1] \to [0,1]$ is a negation function, i.e., a continuous, strictly decreasing function such that $N(0) = 1$ and $N(1) = 0$, and
- $\{\varphi_n\}_{n\geq 2}$ is a standard recursive rule playing a *conjunction* role, in such a way that
$$\varphi_n(a_1, \ldots, a_n) = N^{-1}(\phi_n(N(a_1), \ldots, N(a_n)))$$

Since the particular operator to be applied in each moment is fixed once we know the number of items to be aggregated, such a recursive triplet can be denoted by

$$(\phi, \varphi, N)$$

From such a basic structure, our model should be able to evaluate classification quality, to be analyzed according three aspects:

- Covering: some explanation should be obtained for each object.
- Relevance: each class included in our classification system should be able to explain something about some object.
- Redundancy: each family should not fully overlap with any family of classes.

As pointed out above, we can not impose a perfect classification system. Therefore, we need to build up a measure for each one of these aspects, so the higher degree of relevance, the better; the lower degree of redundancy, the better; the higher degree of explanation, the better (at least as a quite general rule: take into account, for example, that a good classification system, in order to be applied by a decision maker, needs to be understood by this decision maker: if improving mathematical behaviour brings out new classes with no meaning for such a decision maker, it may be the case that such a theoretical improvement is not worth it).

Anyway, behavior according to these three features (covering, relevance and redundancy) should allow some hints about possible improvements to be tried (defining new classes, suppressing some of them or mixing them). In this way, *learning* is justified.

Lets develop here a first proposal showing how covering, relevance and redundancy can be modeled within our aggregative model.

3.1 Covering.

First of all, degree of explanation of each pixel should in principle be as high as possible. Such a degree of explanation can be modeled in terms of

$$\phi\{\mu_c(x)/c \in \mathcal{C}\}$$

Any useful classification system should verify, at least, that

$$\forall x \in X \Rightarrow \exists c \in \mathcal{C}/\mu_c(x) > 0$$

but better if the aggregated value

$$\phi\{\mu_c(x)/c \in \mathcal{C}\}$$

is high enough for each object $x \in X$.

Of course, there is an underlying problem of statistical nature, to be previously solved: highly correlated classes may mean that a class is in fact duplicated, i.e.,

$$\mu_c(x) \simeq \mu_k(x), \qquad \forall x \in X$$

where c, k are two different classes ($c \neq k$). More in general, a class being too close to a linear combination of other classes should be a direct candidate for being deleted from our model. Perhaps a pure statistical analysis in needed at this very first stage. Anyhow, the possible existence of such a bad behavior can be also suggested from a careful analysis of relevance and redundancy, as defined below.

A standard action when covering is too low is a search for some additional classes.

3.2 Relevance.

A *nice* classification system, in general, assumes that no class can be deleted without loosing essential information. As far as explanation does not change when a class is taken away from the model, such a class is less relevant. On the contrary, whenever the degree of explanation goes down significatively for at least one of the objects to be classified, we should consider that such a class is relevant. For example, an *empty* class c ($\mu_c(x) = 0, \forall x \in X$) should be deleted from our model (we are in principle faced again to another statistical problem: how to determine if all those values are significatively close to zero). Thiele [27, 28] excludes from his model those empty classes.

Being a class c relevant means not only that $\mu_c(x) > 0$ for some object x. Of course we expect that our classification system holds that

$$\forall c \in \mathcal{C} \Rightarrow \exists x \in X/\mu_c(x) > 0$$

Again, relevance is also subject to gradation, and it can be measured by means of

$$\phi\{\mu_c(x)/x \in X\}$$

In some way we could think that a class c is less *important* when such a degree of relevance is lower. But it should be taken into account that:

- $\mu_c(x)$ may be even extremely low, but still the only available information about object x (this is the case when $\mu_c(x) > 0$ and $\mu_k(x) = 0, \forall k \neq c$).
- $\mu_c(x)$ may be high, but still not allowing any useful information for classification (for example, if $\mu_c(x) = \mu_c(y), \forall y \in X$).
- $\mu_c(x)$ may be high, but still not the best (object is better described by other clases: for example, when $\mu_k(x) > \mu_c(x)$ but $\mu_k(y) = \mu_c(y), \forall y \neq x$.

Hence, we should analyze relevance of class c by comparing

$$\phi\{\mu_c(x)/\, c \in \mathcal{C}\}$$

and

$$\phi\{\mu_k(x)/\, k \in \mathcal{C}, k \neq c\}$$

for each object x. We shall be accepting relevance of class c whenever there is a *significative* decrease from the first value to the second.

Relevance should also be evaluated for families of alternatives $\mathcal{A} \subset \mathcal{C}$, and not only isolated alternatives $c \in \mathcal{C}$. It implies to compare

$$\phi_n\{\mu_c(x)/\, c \in \mathcal{C}\}$$

and

$$\phi_n\{\mu_k(x)/\, k \in \mathcal{C}, k \notin \mathcal{A}\}$$

for every $A \subset \mathcal{C}, A \neq \emptyset$ (for each object $x \in X$).

This approach supports an alternative search for a smaller family of classes that keeps main explanatory properties (notice that we do not expect to get an automatic classification but a supervised one being interactive with decision makers). A family of classes should be in principle removed from our model if relevance is not appropriate (each element in a family of classes can be not relevant meanwhile the family itself is of course relevant). Nevertheless, the key issue is how classes, and which classes, can be mixed or modified.

Anyhow, as already pointed out, decision maker may be willing to accept that objects belong to several classes, at least at first stages.

A standard action when relevance is too low is to remove some class from the model (or an appropriate redefinition of our classes).

3.3 Redundancy.

Once our family of classes has been reduced taking away an irrelevant enough subset of classes (perhaps classes have been properly redefined), we can assume that each class preserves useful information. But these classes can still overlap in the sense that the same information appears in different classes.

Many classification systems have in mind an ideal *orthogonal* representation, obtained by subsequent modifications in a family of classes till it is suggested that classes do not interact. But this search for a *compact* classification system should keep a real meaning in the decision maker's mind: without a meaning, decision maker can find serious difficulties in managing information. This is an important issue in order to modify or create new classes.

In the same way relevance was analyzed by means of our disjunctive rule ϕ, redundancy can be analyzed by means of our conjunctive rule, φ. For example,

$$\varphi\{\mu_c(x), \mu_k(x)\}$$

can be understood as the degree of overlapping between classes $c, k \in \mathcal{C}, c \neq k$, relative to object x.

In general we may be willing to work with a classification system where classes show high degree of relevance but low degree of redundancy. But we can not expect to be able to define such a nice classification system from the beginning. Covering, relevance and redundancy should play a key role in th subsequent and necessary learning process, giving hints about how to classes should be redefined.

A standard action when redundancy is too high is an appropriate redefinition of our classes (perhaps removing some class from the model). In this sense, redundancy can be sometimes confused with relevance, since actions may be similar. But notice that redundancy focuses on the possibility of excessive (replicated) information, while relevance focuses on unuseful (void) information.

Ruspini's fuzzy partition will appear when ϕ_2 is the Luckasievich t-conorm, and the rule is obtained by associativity, showing complete covering and no redundancy (absolute non relevance is avoided just taking away empty classes).

4 Statistical analysis of a real image.

In order to develop a real application, we consider the image already analyzed in [5] and also presented in [3]: an particular orthoimage in Sevilla province (south Spain) taken on August 18, 1987, by the LANDSAT 5 satellite (*Worldwide Reference System (WRS)* image $202 - 34 - 4$). This image was taken with the Thematic Mapper sensor, which has a spatial resolution of 30 meters. The main objective in [5] was a first fuzzy classification of landscape by means of an unsupervised algorithm initially presented in Amo *et al.* [2]. Fuzzy classes obtained in [5] were compared with those obtained from the $ERDAS^{TM}$ program on the same image, showing the accuracy of our fuzzy approach. As a conclusion, it was pointed out that whenever real classes are natural fuzzy classes, with gradual transition between classes, then its fuzzy representation will be more easily understood -and therefore accepted- by users. In now propose to consider standard statistical methods in order to *quantify* the quality of the classification obtained in [5]. In this way we can focus our future attention on the search for an alternative improved family of classes (see [14] and [24] for a more general discussion on fuzzy methods in remote sensing end ecological modeling).

Hence, in this section we shall consider the solution shown in [5] and check some statistical properties of its associated covering, relevance and redundancy indexes.

Results below refer to the following disjunctive and conjunctive rules:

- $\phi(a_1 \ldots a_n) = \frac{1-\prod_{k=1,n}(1-a_k)}{1+2\prod_{k=1,n}(1-a_k)}$

- $\varphi(a_1, \ldots a_n) = \frac{3\prod_{k=1,n} a_k}{1+2\prod_{k=1,n} a_k}$

where parameters have been fixed after some experiments (see [1]). Statistical tests will be applied assuming that pixels represent a random sample, in such a way we can in some way check the percentage of pixels being covered by the classification system proposed in [5], or the percentage of pixels showing relevance or redundancy for that classification system.

4.1 Covering.

First thing to be checked is the global quality of the selected family of classes.

In order to measure covering, we propose to consider some statistical tests

$$H_0 : \phi\{\mu_c, c \in \mathcal{C}\} \leq \alpha$$
$$H_1 : \phi\{\mu_c, c \in \mathcal{C}\} > \alpha$$

Value α appears as a quality level, in such a way that rejecting H_0 means that at least $100\alpha\%$ data is being explained.

Then we can take as a covering measure the maximum value α_0 such that the above test rejects H_0, for all $\alpha \leq \alpha_0$.

High values of α_0 will be associated to high covering. It will depend on our particular experience to decide if a certain level suggests high enough covering.

When classical non parametric Wilcoxon tests were applied to the above referenced Sevilla orthoimage data (see [5]), the algorithm proposed in [2] for two classes gave a too low covering index. In fact, as pointed out in [5], with two classes we were not able to recognize key objects in the picture. So, we considered three classes.

Classification algorithm in [2] was then developed now imposing three classes. If we just denote these three classes by μ_1, μ_2 and μ_3, then the application of Wilcoxon tests to

$$H_0 : \phi(\mu_1, \mu_2, \mu_3) \leq \alpha$$
$$H_1 : \phi(\mu_1, \mu_2, \mu_3) > \alpha$$

gives $\alpha_0 = 0.6$, which we considered good enough (analogous results were obtained for other non parametric tests).

A careful analysis pixel by pixel may help to guess how classes can be redefined in order increase covering.

4.2 Relevance

Once we have an explanatory enough family of classes, we trie to check at what extent it is close to a true partition, i.e., at what extent classes are relevant and non redundant.

As already pointed out, once we obtained three fuzzy classes in [5], we should be comparing the aggregated explanatory index for those three classes

$$\phi(\mu_1, \mu_2, \mu_3)$$

with the analogous explanatory index for every subset of classes:

$$\phi(\mu_2,\mu_3),\phi(\mu_1,\mu_3),\phi(\mu_1,\mu_2)$$

at a first stage. In this way we can evaluate relevance of isolated classes μ_1, μ_2, μ_3.

Whenever

$$\phi(\mu_1,\mu_2,\mu_3)$$

is significatively higher than $\phi(\mu_2,\mu_3)$,$\phi(\mu_1,\mu_3)$ or $\phi(\mu_1,\mu_2)$ it will be respectively suggesting that class μ_1,μ_2 or μ_3 is relevant. Non relevant classes should be in general suppressed, but we rather expect that a careful analysis pixel by pixel will help us to guess how classes can be redefined in order to avoid irrelevant information.

If we fix a minimum level for each difference,

$$\phi(\mu_1,\mu_2,\mu_3)-\phi(\mu_i,\mu_j)>\alpha$$

in order to suggest relevance, we could consider some appropriate test in order to decide if such a level can be assumed. A relevance measure can be defined by means of the maximum value α_0 such that we reject H_0 for all $\alpha \leq \alpha_0$ in a test

$$H_0:\phi(\mu_1,\mu_2,\mu_3)-\phi(\mu_i,\mu_j)\leq\alpha$$
$$H_1:\phi(\mu_1,\mu_2,\mu_3)-\phi(\mu_i,\mu_j)>\alpha$$

for each pair $\{\mu_i,\mu_j\}$ of classes, $i \neq j$.

High values of α_0 will suggest relevance of isolated class. Results obtained for our Sevilla image were the following:

- For class 1, comparison between $\phi(\mu_1,\mu_2,\mu_3)$ and $\phi(\mu_2,\mu_3)$ gave $\alpha_0 = 0.14$, meaning a 14% information directly associated to class 1.
- For class 2, comparison between $\phi(\mu_1,\mu_2,\mu_3)$ and $\phi(\mu_1,\mu_3)$ gave $\alpha_0 = 0.15$.
- For class 3 comparison between $\phi(\mu_1,\mu_2,\mu_3)$ and $\phi(\mu_1,\mu_2)$ lead to $\alpha_0 = 0.13$.

All these values were considered high enough. Hence, we can assume that each isolated class within the three classes solution proposed in [5] is relevant, since suppressing from the model any of them the explanatory level would decrease too much.

4.3 Redundancy

Overlapping in our example [5] can be checked by means of

$$\varphi(\mu_i,\mu_j)$$

Too much redundancy will in general suggest deleting some class, but we again rather expect to be able to guess how classes should be redefined.

Following the above approach, we can fix a minimum level of redundancy α, and then evaluate the maximum value α_0 such that the test

$$H_0 : \varphi(\mu_i, \mu_j) \leq \alpha$$
$$H_1 : \varphi(\mu_i, \mu_j) > \alpha$$

concludes rejection for every $\alpha \leq \alpha_0$. A high value α_0 will suggest overlapping. A careful analysis pixel by pixel may help of to guess how classes can be redefined in order to avoid redundancy.

In our example, again for three classes and applying the Wilcoxon test, we obtain the following results:

- $\varphi(\mu_2, \mu_3)$: $\alpha_0 = 0.42$.
- $\varphi(\mu_1, \mu_2)$: $\alpha_0 = 0.44$.
- $\varphi(\mu_1, \mu_3)$: $\alpha_0 = 0.42$.

At a first stage none of them seem to be too excessive, so in principle we can accept the classification system given by the three fuzzy classes defined in [5]. But of course, depending on the precision we pursue we may be forced to redefine those three classes. A careful study of behavior of each index pixel by pixel will for sure give a hint about the characteristics of the new classification system we are looking for, hopefully reducing redundancy.

5 Final comments.

Future research should implement more complex inner structures and relations between classes. In fact, our recursive model can be strictly applied when those classes show an underlying linear order.

Moreover, although it has been shown that a battery of statistical tests can play an interesting role in order to evaluate the quality of actual classification system, and therefore they will help us to learn about better classification systems, the key issue still is how new classes and alternatives can be created (see Roy [22] and Shafer [26]). We still need much more experiences.

Acknowledgements This research is being supported by the Government of Spain and a Del Amo joint project between Complutense University and the University of California.

References

[1] A. Amo (1999): *Modelos de aggregación recursiva y su aplicación a la clasifiación difusa de imágenes digitales*; Ph.D. Thesis, Complutense University, Madrid (in spanish).

[2] A. del Amo, J. Montero and G. Biging (2000): *Classifying pixels by means of fuzzy relations*; Int. J. General Systems 29:605–621.

[3] A. Amo, J. Montero and G. Biging (2000): *Relevancia y redundancia en sistemas de clasificación difusa*; Congreso Español sobre Tecnologías y Lógica Fuzzy, ESTYLF 2000 (A. Ollero, S. Sánchez, B. Arrue and I. Baturone, eds., Universidad de Sevilla, Sevilla); pp. 521-525 (in spanish).

[4] A. del Amo, J. Montero y E. Molina (2001): *Representation of consistent recursive rules*; European Journal of Operational Research 130:29–53.

[5] A. Amo, J. Montero, A. Fernández, M. López, J. Tordesillas and G. Biging (to appear): *Spectral fuzzy classification: an application*; IEEE Transactions of Systems, Man and Cybernetics.

[6] B. de Baets (1998): *Uninorms: the known classes*; in D. Ruan, H.A. Abderrahim y P. D'hont (eds.), *Fuzzy Logic and Intelligent Technologies for Nuclear Science and Industry* (World Scientific, Singapore), pp. 21–28.

[7] J.C. Bezdek and J.D. Harris (1978): *Fuzzy partitions y Relations: an axiomatic basis for clustering*; Fuzzy Sets and Systems 1:111–127.

[8] J.C. Bezdek (1992): *Fuzzy models for pattern recognition: background, significance and key points*; in J.C. Bezdek and S.K. Pal (eds.), *Fuzzy Models for Pattern Recognition* (IEEE Press, New York), pp. 1–27.

[9] V. Cutello and J. Montero (1999): *Recursive connective rules*; International Journal of Intelligent Systems 14:3–20.

[10] J. Dombi (1982): *Basic concepts for a theory of evaluation: the aggregative operator*; European Journal of Opearational Research 10:282–293.

[11] J. Dombi (1982): *A general class of fuzzy operators, the De Morgan class of fuzzy operators and fuzziness measures induced by fuzzy operators*; Fuzzy Sets and Systems 8:149–163.

[12] J.C. Fodor and M. Roubens (1994): *Valued preference structures*; European Journal of Operational Research 79:277–286.

[13] J.C. Fodor and M. Roubens (1994): *Fuzzy Preference Modelling and Multicriteria Decision Support* (Kluwer Academic Publishers).

[14] G.M. Foody (1999): *The continuum of classification fuzziness in thematic mapping*, Photogrammetric Engineering and Remote Sensing 65:443–451.

[15] T.C. Koopmans (1972): *Representation of preference ordering with independent components of consumption*; in C.B. McGuire and R. Radner (eds.), *Decision and Organization* (North-Holland, Amsterdam), pp. 57–78 (2nd edition by the University of Minnesota Press, 1986).

[16] K.T. Mak (1987): *Coherent continuous systems and the generalized functional equation of associativity*; Mathematics of Operations Research 12:597–625.

[17] J.L. Marichal, P. Mathonet and E. Tousset: Characterization of some aggregation functions stable for positive trasnformations. *Fuzzy Sets and Systems* 102:293–314 (1999).

[18] M. Mas, G. Mayor, J. Suñer and J. Torrens (1998): *Generation of multidimensional aggregation functions*; Mathware and Soft Computing 5:233–242.

[19] S. Medasani, J. Kim and R. Krishnapuram (1998): *n overview of membership function generation techniques for pattern recognition*; Int. J. Approximate Reasoning 19:391–417.

[20] J. Montero, J. Tejada and V. Cutello (1997): *A general model for deriving preference structures from data*; European Journal of Operational Research 98:98–110.

[21] M. Roubens (1978): *Pattern classification problems and fuzzy sets*; Fuzzy Sets and Systems 1:239–253.

[22] B. Roy (1993): *Decision science or decision-aid science*; European Journal of Operational Research 66:184–203.

[23] E.H. Ruspini (1969): *A new approach to clustering*; Information and Control 15:22–32.

[24] A. Salski, O. Fränzle and P. Kandzia (1995): Fuzzy Logic in Ecological Modelling (Special issue); Ecological Modelling 85.

[25] B. Schweizer and A Sklar (1983): *Probabilistic Metric Spaces*; North-Holland, New York.

[26] G. Shafer (1992): *Rejoinders to comments on "Perpectives on the theory and practice of belief functions"*; Int. J. Approximate Reasoning 6:445–480.

[27] H. Thiele (1996): *A Characterization of RUSPINI-Partitions by Similarity Relations*; in *IPMU'96*, pp. 389–394.

[28] H. Thiele (1996): *A Characterization of Arbitrary RUSPINI-Partitions by Fuzzy Similarity Relations*; in *IPMU'96*, pp. 131–134.

[29] R.R. Yager (1993): *Families of OWA operators*; Fuzzy Sets and Systems 59:125–148.

[30] R.R. Yager (1993): *MAM and MOM operators for aggregation*; Information Sciences 69:259–273.

[31] R.R. Yager y A. Rybalov (1996): *Uninorm aggregation operators*; Fuzzy Sets and Systems 80:111–120.

[32] L.A. Zadeh (1965): *Fuzzy Sets*; Information and Control 8:338–353.

[33] L.A. Zadeh (1973): *Outline of a new approach to the analysis of complex systems and decision processes*; IEEE Transactions on Systems, Man and Cybernetics 1:28–44.