Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Operational Research 156 (2004) 495–507
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/dsw

Computing, Artificial Intelligence and Information Technology

Fuzzy classification systems

A. Amo [a], J. Montero [a,*], G. Biging [b], V. Cutello [c]

[a] Faculty of Mathematics, Department of Statistics & O.R., Complutense University, Madrid, 28040, Spain
[b] Environmental Sciences, University of California, Berkeley, CA, USA
[c] Department of Mathematics, Catania University, Catania, Italy

Received 6 December 2000; accepted 4 December 2002

Abstract

In this paper it is pointed out that a classification is always made taking into account all the available classes, i.e., by means of a *classification system*. The approach presented in this paper generalizes the classical definition of fuzzy partition as defined by Ruspini, which is now conceived as a quite often desirable objective that can be usually obtained only after a long learning process. In addition, our model allows the evaluation of the resulting classification, according to several indexes related to covering, relevance and overlapping.

Keywords: Classification; Fuzzy sets; Fuzzy partition; Multicriteria analysis

1. Introduction

Since the beginning of the history of Fuzzy Sets Theory [37], Classification Models and Control Theory have been two central fields for their theoretical and practical developments (see, e.g., [7,38]). In fact, many problems within both fields are naturally formalized by introducing fuzzy concepts. In some cases, a fuzzy approach seems to offer a useful simplification of a too complex reality. This is mainly the case for control problems. In other cases, the concepts users have in mind are fuzzy in nature, in the sense that they allow degrees of verification. This is the case in many classification problems, where the introduction of crisp classes represent an unrealistic oversimplification of reality, leading to obviously wrong interpretations, when compared to a direct observation.

In many fuzzy classification applications, a set of classes $\mathscr{C}$ is assumed. Such a set of classes can be obtained in may different ways. The first objective, once such a family of available classes has been defined, is to determine the degree $\mu_c(x)$ to which object x belongs to class $c \in \mathscr{C}$, for every object x under consideration, $x \in X$. In this way, it is being defined a membership function

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

for each class $c \in \mathscr{C}$ (see, e.g., [26]).

It can be thought that each membership function can be evaluated by itself, without taking into account the remaining classes. But it is a fact that most users will find serious difficulties in assigning degrees of membership to one class without taking into consideration the remaining possibilities for

* Corresponding author. Tel.: +34913944522; fax: +34913944607.
E-mail address: javier_montero@mat.ucm.es (J. Montero).

0377-2217/$ - see front matter
doi:10.1016/S0377-2217(03)00002-X

classification. Classification procedures usually require a previous look at all available classes. Hence, classification methods are in general highly dependent on the family of classes the user is forced to consider (even in a crisp context, users frequently have a look at all possible choices before choosing a particular class for a given object). Assigned degrees of membership do have influence on other degrees of membership yet to be assigned.

Therefore, a key concept in classification is the notion of partition, since it produces a structured family of classes. Each class is strongly related to each other, showing a specific structure (in a crisp context, for example, decision maker is being forced to choose one and only one class for each object).

Fuzzy partitions were introduced by Ruspini [28] in 1969 (see also [7,8,14,18]): given a discrete family $\mathscr{C}$ of *classes*, it is assumed that for every object under consideration

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

always holds ($\forall x \in X$). Each object x may belong to several classes -to certain degrees-, and the total degree of membership is distributed among all classes. In this way, the classical crisp partition concept was generalized. Indeed, whenever

$$\mu_c(x) \in \{0,1\} \quad \forall x,\ \forall c$$

holds, each object will be in one and only one class: for all $x \in X$ there exist $c \in \mathscr{C}$ such that $\mu_c(x) = 1$ and hence $\mu_k(x) = 0$, for all $k \neq c$.

From our point of view, Ruspini's proposal may in principle represent a desirable situation, in the sense that it seems to model a typical situation were every object seems to be explained, in some way, with the minimum amount of information. But again, it seems to be a too restrictive definition within fuzzy modeling. In most cases, the fuzzy classes under consideration do not verify the conditions of such a Ruspini fuzzy partition (the decision maker will not meet those requirement if not artificially forced). Perhaps it is is only after a long learning process that users are able to get a new family of fuzzy classes, assuring that every object is fully explained without superfluous information. Classes are defined in advance to membership assignments, and therefore no condition can be naturally imposed on them.

Moreover, such an ideal Ruspini's classification system may be not possible, or even not desirable, when faced with some particular problems. Some fruit classification problems, for example, require (due to market restrictions) a large number of classes, and each piece of fruit is allowed to be simultaneously associated with several different classes.

In a complex classification problem, as in Remote Sensing for example (see [2] for a discussion on a particular algorithm and its application), we have to classify pixels in a picture representing landscape that have been obtained from a satellite (see also [17] for a general discussion). Each pixel in a digital image represents several kilometers or in the best case meters. If our classification system is a crisp partition, we should decide one and only one class for each pixel, something very difficult to accept since it is obvious that many pixels may not be homogeneous in their internal description. A fuzzy approach may in principle show clear advantages, since it allows a natural representation of mixtures and transition zones (most earth surface classes do not present sharp boundaries). But if a Ruspini partition is taken as our classification system, we are forced to calibrate degrees of membership in such a way their sum throughout all classes is exactly one. This is also a strong restriction in practice, not only because of standard measure difficulties, but because the available family of classes should allow such a situation. The standard procedure is to propose (perhaps on the basis of personal experience) a particular family of *natural* classes, and then check degrees of membership. Obviously, we cannot expect that the classes we have proposed at a first stage are just the perfect ones we were seeking for our particular purposes. Most probably, we shall realize that a better set of classes is needed. It is not trivial at all to get a Ruspini's partition as a first stage. Although some practical difficulties of Ruspini's partitions can be partially overcome by a weaker approach proposed by some other authors (see, e.g., [31,32]), it is clear that Ruspini's classification system cannot always be used (see example in [1] about a general computer security problem).

In this paper we propose to analyze classification systems by means of aggregative models (see, e.g., [12,13,24]), which should present Ruspini's partition as a particular additive solution. Then we shall take into account some results already obtained in other formal contexts in order to propose a more general classification structure (see [15,16, 24] but also [4,10]). In particular, we shall generalize the approach presented in [14,18] and in particular [1], based upon De Morgan triples.

2. Aggregation processes and fuzzy classification systems

In a fuzzy classification process, one of the main goals is to amalgamate the information we have obtained about each object. If such information is given by means of degrees of membership, we can find in the literature that additive rules [28] and some t-conorms (see, e.g., [29]) are useful for forming conjunctive rules. Information is in this way aggregated into one single index. Since we usually do not know in advance the number of chunks of information we should aggregate, classical approaches assume the existence of a basic binary operator being associative, in such a way that a sequential application of such an operator will give us the aggregated information, no matter the dimension of information. But we know that not every operator is associative (OWA operators [34], for example, where it is necessary to know in advance the dimension of the data set). Otherwise we may have problems in defining what a rule is (see [34] for a possible solution for OWA operators).

In this paper we consider the recursive approach introduced in [10], which generalizes the concept of associativity, and the particular representation results obtained in [4].

Recursiveness is a property of a sequence of operators $\{\phi_n\}_{n>2}$ allowing the aggregation of any number of items: ϕ_2 tells us how to aggregate two items, ϕ_3 tells how to aggregate three items and so on, in such a way that being faced with n items we shall aggregate them by means of ϕ_n. Recursiveness assures consistency of such a family of operators by assuming that such a rule is *operational*, in the sense that it can be evaluated both from left and right by means of a sequence of binary operators. In general, each new data implies a modification of the binary operator, in such a way that these binary operators *evolve*.

By definition, we assume that the aggregation of only one item is always the identity.

Definition 2.1. A *recursive rule* ϕ is a family of aggregation functions

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

such that there exist an ordering rule π and two sequences of binary operators

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

and

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

such that

$$\begin{aligned}\phi_n(\pi(x_1),\ldots,\pi(x_n)) &= L_n(\phi_{n-1}(\pi(x_1),\ldots,\pi(x_{n-1})),\pi(x_n))\\ &= R_n(\pi(x_1),\phi_{n-1}(\pi(x_2),\ldots,\pi(x_n))).\end{aligned}$$

In other words, a recursive rule is a family of operators allowing a sequential reckoning by means of a successive application of binary operators, once data have been properly ordered: the ordering rule assures that new data do not introduce modifications in the relative position of items already ordered (see [4,10]). Of course, not every sequence of binary operators will define a recursive rule. Recursiveness is assuring a certain consistency on that sequence of binary operators, in such a way that consistency is in fact assured, and it can be properly understood as a rule.

In this paper we shall assume, by definition, that our recursive rules can be represented by means of binary operators $\{L_n\}_{n>1}$ and $\{R_n\}_{n>1}$ being all these binary operators continuous and nondecreasing functions such that

$$L_n(0,0) = R_n(0,0) = 0 \quad \forall n$$

and

$$L_n(1,1) = R_n(1,1) = 1 \quad \forall n.$$

A *standard* recursive rule will be one based upon the identity ordering rule (i.e., the ordering rule that keeps the order of data as they are given to us). A *commutative* recursive rule will appear when the same results follows no matter the ordering rule we consider (see again [4,10]).

Obviously, in case such a sequence of binary operators is given by a unique binary operator ($L_i = R_j, \forall i, j$), we should be talking about a rule based upon an associative binary operator. Hence, recursiveness indeed allows the generalization of associativity (an associative rule is a recursive rule based upon a unique binary operator).

Since we are in principle assuming that the family of classes $\mathscr{C}$ is not definitive, our classification system should show a certain ability to detect main classification problems, i.e., if classes overlap and some of them are even redundant or if an extra class is needed. Hence, our classification system must contain not only the family of classes $\mathscr{C}$ under consideration at a certain stage, but some tools in order to evaluate standard characteristics of a classification, like covering, relevancy and redundancy. Goodness of the resulting classification should be analyzed taking into account measures related to these concepts. The basic structure we propose below will allow it, assuming arbitrary recursive rules instead of associative rules in a classical De Morgan's triple (see [1] but also [14,18]).

Definition 2.2. Let us assume a finite set of objects X. A *fuzzy classification system* is a countable family of fuzzy classes $\mathscr{C}$ (each $c \in \mathscr{C}$ with its associated membership function $\mu_c : X \to [0,1]$), together with a *recursive triplet* (ϕ, φ, N) where

- ϕ is a standard recursive rule such that $\phi_2(0,1) = \phi_2(1,0) = 0$,
- $N : [0,1] \to [0,1]$ is a strict negation function (see [33]), i.e., a bijective strictly decreasing function such that $N \circ N^{-1}(x) = x$ for all $x \in [0,1]$, and
- φ is another standard recursive rule such that

$$\varphi_n(x_1, x_2, \ldots, x_n) = N^{-1}[\phi_n(N(x_1), N(x_2), \ldots, N(x_n))] \quad \forall n > 1.$$

A *fuzzy classification system* can be therefore denoted by

$$(\mathscr{C}; \phi, \varphi, N).$$

First of all, notice that φ is a *disjunctive recursive rule*, in the sense that $\varphi_n(x_1, x_2, \ldots, x_n) = 1$ whenever $\exists j / x_j = 1$. As a direct consequence, ϕ is a *conjunctive recursive rule*, in the sense that $\phi_n(x_1, x_2, \ldots, x_n) = 0$ whenever $\exists j / x_j = 0$.

Theorem 2.1. *Let us consider a fuzzy classification systems $(\mathscr{C}; \phi, \varphi, N)$ defined as above. Then:*

1. *If $x_j = 1$ for some j, then $\varphi_n(x_1, x_2, \ldots, x_n) = 1$.*
2. *If $x_j = 0$ for some j, then $\phi_n(x_1, x_2, \ldots, x_n) = 0$.*

Proof. We shall prove only the first assertion, by induction, and the second assertion has a similar proof: in fact, if we consider $n = 3$ and there exists $x_1 = 1$ or $x_2 = 1$ or $x_3 = 1$, then either $\varphi_2(x_1, x_2) = 1$ or $\varphi_2(x_2, x_3) = 1$ holds, in such a way that either

$$\begin{aligned}\varphi_3(x_1, x_2, x_3) &= L_3(\varphi_2(x_1, x_2), x_3) = \varphi_3(1, 1, x_3) \\ &= R_3(1, \varphi_2(1, x_3)) = R_3(1, 1) = 1\end{aligned}$$

or

$$\begin{aligned}\varphi_3(x_1, x_2, x_3) &= R_3(x_1, \varphi_2(x_2, x_3)) = \varphi_3(x_1, 1, 1) \\ &= L_3(\varphi_2(x_1, 1), 1) = L_3(1, 1) = 1.\end{aligned}$$

Now, if we assume as induction hypothesis that $\varphi_n(y_1, \ldots, y_n) = 1$ whenever $\exists j / y_j = 1$ for a fixed n, then either $\varphi_n(x_1, \ldots, x_n) = 1$ or $\varphi_n(x_2, \ldots, x_{n+1}) = 1$, and hence either

$$\begin{aligned}\varphi_{n+1}(x_1, \ldots, x_{n+1}) &= L_{n+1}(\varphi_n(x_1, \ldots, x_n), x_{n+1}) \\ &= R_{n+1}(1, \varphi_n(1, \ldots, 1, x_3)) \\ &= R_{n+1}(1, 1) = 1\end{aligned}$$

or

$$\begin{aligned}\varphi_{n+1}(x_1, \ldots, x_{n+1}) &= R_{n+1}(x_1, \varphi_n(x_2, \ldots, x_{n+1})) \\ &= L_{n+1}(\varphi_n(x_1, 1, \ldots, 1), 1) \\ &= L_{n+1}(1, 1) = 1\end{aligned}$$

and in any case $\varphi_{n+1}(x_1, \ldots, x_{n+1}) = 1$ holds. □

Moreover, it can be checked that ϕ being a standard recursive rule, then φ is in fact assured to be a standard recursive rule also, as shown in the following theorem.

Theorem 2.2. *Let $\{\phi_n\}_{n>1}$ be a standard recursive connective rule and $N : [0,1] \to [0,1]$ a strict negation function. Then*

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

such that

$$\varphi_n(x_1,\dots,x_n) = N^{-1}(\phi_n(N(x_1),\dots,N(x_n))) \quad \forall (x_1,\dots,x_n) \in [0,1]^n$$

is another standard recursive connective rule.

Proof. From definition of $\{\phi_n\}_{n>1}$ we know that there exist a sequence of binary operators

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

such that for all $(x_1,\dots,x_n) \in [0,1]^n$, $\forall n \geqslant 3$ we have

$$\phi_n(x_1,\dots,x_n) = L_n(\phi_{n-1}(x_1,\dots,x_{n-1}),x_n) = R_n(x_1,\phi_{n-1}(x_2,\dots,x_n)).$$

Therefore, for all $n \geqslant 3$,

$$\begin{aligned}\varphi_n(x_1,\dots,x_n) &= N^{-1}(\phi_n(N(x_1),\dots,N(x_n)))\\ &= N^{-1}(L_n(\phi_{n-1}(N(x_1),\dots,N(x_{n-1})),N(x_n)))\\ &= N^{-1}(R_n(N(x_1),\phi_{n-1}(N(x_2),\dots,N(x_n)))).\end{aligned}$$

Then, just defining

$$I_n(a,b) = N^{-1}(L_n(N(a),N(b)))$$

for all n, we can check that φ_n is left recursive:

$$\begin{aligned}\varphi_n(x_1,\dots,x_n) &= N^{-1}(\phi_n(N(x_1),\dots,N(x_n)))\\ &= N^{-1}(L_n(\phi_{n-1}(N(x_1),\dots,N(x_{n-1})),N(x_n)))\\ &= N^{-1}(L_n(N\circ N^{-1}(\phi_{n-1}(N(x_1),\dots,\\ &\qquad N(x_{n-1})),N(x_n))))\\ &= N^{-1}(L_n(N(\varphi_{n-1}(x_1,\dots,x_{n-1})),N(x_n)))\\ &= I_n(\varphi_{n-1}(x_1,\dots,x_{n-1}),x_n).\end{aligned}$$

Analogously, right recursiveness follows taking $D_n(a,b) = N^{-1}(R_n(N(a),N(b)))$, in such a way that

$$\varphi_n(a_1,\dots,a_n) = D_n(a_1,\varphi_{n-1}(a_2,\dots,a_n))$$

for all $(a_1,\dots,a_n) \in [0,1]^n$. □

3. Covering

Then, from the degree to which an object $x \in X$ belongs to each class $c \in \mathscr{C}$ we can obtain an aggregated value (depending on the particular *disjunctive* recursive rule we have chosen)

$$\mu_{\mathscr{C}}(x) = \varphi\{\mu_c(x)/c \in \mathscr{C}\}$$

which can be understood as the degree to which such an object $x \in X$ is *explained* by such a family of classes $\mathscr{C}$. The higher all these values $\mu_{\mathscr{C}}(x)$, the better. In case $\mu_{\mathscr{C}}(x) = 1$ for all $x \in X$, it can be understood that the whole family of objects is fully *covered* by our family of fuzzy classes $\mathscr{C}$, being a trivial generalization of classical approaches, including [18]. The lower such an aggregated covering value $\mu_{\mathscr{C}}(x)$, the stronger claim for a new class.

But it is obvious that such a numerical analysis may lead to trivial paradoxes, for example by replicating several times a given class. The family $\mathscr{C}$ of classes should of course *explain* as much as possible variation in all the objects (such a family of classes must represent a nice *covering* of objects), but it should be also as compact as possible, taking into account not only *relevant* classes, but also reducing possible overlapping information (*redundancy*). Let us discuss these two extra notions (relevancy and redundancy) in the next sections.

4. Relevancy

As already pointed out, in order to assure we are in principle capturing all the information we need for classification, we are in principle interested in families of classes $\mathscr{C}$ fully covering our set of objects, in the sense that

$$\varphi\{\mu_c(x)/c \in \mathscr{C}\} = 1 \quad \forall x \in X$$

or at least *close enough* to 1 for each object. The value

$$\varphi\{\mu_c(x)/c \in \mathscr{C}\}$$

can indeed be considered a measure of how close we are to the desired full covering (with respect to object $x \in X$).

But it has been already made clear that replicated classes (i.e., $c, d \in \mathscr{C}$, $c \neq d$, such that $\mu_c(x) = \mu_d(x), \forall x \in X$) should be avoided in our model. The same applies to any *void* class c (i.e., a class $c \in \mathscr{C}$ such that $\mu_c(x) = 0, \forall x \in X$). These two extreme situations can be easily solved, by just deleting some classes after a trivial comparison. The difficulty will appear when we find out that we are *close* to some of those two extreme cases. In both cases, a class is *almost* useless when it barely helps to explain anything (at least when applied to our particular fixed set of objects). These two problems can be in principle addressed by pure statistical techniques, but it should be pointed out that it may happen that a certain class is showing statistical nonsignificance but still that information may be the only hint we have about how to classify some pixels.

In Thiele [31,32], for example, void classes are excluded axiomatically. Certainly, we should take into account in our model only those classes $c \in \mathscr{C}$ being *relevant* in the sense that $\mu_c(x) > 0$ for some object $x \in X$. But being *relevant* is also a matter of degree, and it is also relative to the other classes in $\mathscr{C}$.

A first proposal for measuring *relevancy* of class $c \in \mathscr{C}$ is to evaluate

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

in such a way that class c can in principle be assumed less important as far as such a value is low. But such an approach may be misleading:

- $\mu_c(x)$ may be low, but still the only information we have about object x. This is the case when $\mu_c(x) > 0$ and $\mu_d(x) = 0, \forall d \neq c$.
- $\mu_c(x)$ may be high, but still that object x is much better described by other classes. For example, if $\mu_d(x) > \mu_c(x)$ and $\mu_d(y) = \mu_c(y), \forall y \neq x$.
- $\mu_c(x)$ may be high, but still may not give us any discriminant information about object x. For example, if $\mu_c(x) = \mu_c(y), \forall y \neq x$.

The above examples give a hint of some difficulties an alternative pure statistical analysis may present, although it is clear that statistical support is a must in order to clarify how relevant a class is.

Anyhow, *relevancy* should always be evaluated not for each isolated class c being a candidate *to be deleted*, but for a selected set of classes $\mathscr{A}$.

In order to analyze *relevancy*, we propose to compare behaviour of each nonempty class $\mathscr{A}$ with behaviour of the remaining $\mathscr{C} - \mathscr{A}$ classes being kept in the model, taking into account

(1) $\varphi\{\mu_c(x)/c \in \mathscr{C}\}$,
(2) $\varphi\{\mu_k(x)/k \in \mathscr{A}\}$,
(3) $\varphi\{\mu_k(x)/k \in \mathscr{C} - \mathscr{A}\}$,

for every object $x \in X$, being $\mathscr{A}$ a fixed nonempty family of classes, $\mathscr{A} \subset \mathscr{C}$. For example, when (1) is significantly higher than (3), then we can in principle suggest that $\mathscr{A}$ is a relevant family of classes, even if (2) is not high. And whenever (1) is nonsignificantly different than (3), we can in principle suggest that $\mathscr{A}$ is a nonrelevant family of classes, even when (2) is not low.

A complete track of relevancy may therefore imply a large number of aggregations, but evaluating relevancy of isolated classes may be enough for most practical purposes. Additional relevancy indexes will most probably be taken into account in a second stage, only if we are willing a more accurate classification, in case results show some deficiency to the decision maker. Notice that it may happen that a relevant family of classes contains no relevant class.

Anyhow, our approach represents the basis for a formal analysis of *relevancy*, searching for the smallest set of classes which maintains the user's desired explanatory properties.

The *relevancy* issue can be therefore addressed as a *dimensionality reduction* problem, in such a way that other statistical and nonstatistical representation models can be introduced (see, e.g., [25,36]). Obviously, a class which gives no additional information at all about how to classify our objects does not deserve to be kept in the model. Again, the key problem is to decide how many and which ones of those classes can be deleted while still keeping enough explanatory power. In practice we shall always look for an *appropriate* number of classes (the lower the number of classes, the better) to maximize classification accuracy. But this is a decision that should be left in the hands of the decision maker.

Notice that the user may be willing to accept that objects belong to several classes, even a clear crisp overlapping (i.e., $\mu_c(x) = \mu_k(x) = 1$ for some $c \neq k$), if such a situation is considered by the user as *relevant* for the particular classification problem. Such a *relevancy* issue should not be confused with the *redundancy* issue, to be discussed in the next section. In fact, the goals are different, mathematical treatments are different, and both problems are addressed at different stages.

5. Redundancy

Once the initial set of classes has been analyzed and *nonrelevant* classes have been suppressed from the model, we can assume that every class is giving us some amount of useful information. But classes may still overlap.

From a pure representation point of view, the less overlapping, the better. This *redundancy* property refers to a certain *orthogonality* of the family of classes, which is then viewed as a particular *representation system* of the set of objects. *Redundancy* suggests the possible existence of an alternative representation, to be found by means of an appropriate redefinition of our family of classes. This is indeed a key issue, related to that *path of constructivism* claimed by Roy [27] in a multicriteria decision making context (see also [6,30]). It must be taken into account how difficult is to redefine previous classes (combining already known observable characteristics) and how difficult is to define extra classes (by means perhaps of some characteristic not previously taken into account). Moreover, it is very important that classes have some real meaning for users (otherwise it may be extremely difficult to assign degrees of membership and allow decision maker the possibility of checking the whole procedure).

Once *relevancy* has been studied at a first stage (on the basis of our φ disjunction rule), *class overlapping* will be studied by means our conjunction rule ϕ. In fact, the value

$$\phi\{\mu_c(x), \mu_d(x)\}$$

can be understood as the degree of overlapping between classes $c, d \in \mathscr{C}$ with respect to object $x \in X$.

In particular, when dealing with crisp partitions ($\mu_c(x) \in \{0, 1\}, \forall x, \forall c$), for each object x it is assumed the existence of a unique class c each object x belongs to, in such a way that $\mu_c(x) = 1$ for some c and $\mu_d(x) = 0$ for all $d \neq c$. So,

- $\varphi\{\mu_c(x)/c \in \mathscr{C}\} = 1 \quad \forall x \in X$, and
- $\phi\{\mu_c(x), \mu_d(x)\} = 0 \quad \forall x \in X,\ \forall c \neq d$.

Analogously, when dealing with Ruspini partitions [28], since

$$\sum_{c \in \mathscr{C}} \mu_c(x) = 1 \quad \forall x \in X$$

we have that choosing the Luckasievich's t-conorm S and t-norm T as our disjunctive rule and conjunctive rule, respectively, we can assure that (see [8,14,18])

- $S\{\mu_c(x)/c \in \mathscr{C}\} = \min\{\sum_{c \in \mathscr{C}} \mu_c(x), 1\} = 1$ $\forall x \in X$,
- $T\{\mu_c(x), \mu_d(x)\} = \max\{0, \mu_c(x) + \mu_k(x) - 1\} = 0$ $\forall x \in X,\ \forall c \neq d$.

Overlapping can of course be evaluated not only for pairs of classes, but for every subset of classes $\mathscr{A} \subset \mathscr{C}$ with at least two elements, allowing the possibility of a better insight into the structure and relationships among those classes:

$$\phi\{\mu_c(x)/c \in \mathscr{A}\}.$$

Again, we expect that redundancy for pairs of classes will be enough in most practical problems, at least at a first stage. Additional redundancy indexes will be evaluated only in case results suggest decision maker to trie a more accurate classification system.

6. Fuzzy partition systems

A *fuzzy classification system* $(\mathscr{C}; \phi, \varphi, N)$ will be meaningless in some cases, whenever it does not allow any discrimination among objects. This may happen either because no explanation is attained, or because all classes fully overlap. Classification systems should be able to explain something about each object, and no classes should fully overlap.

Definition 6.1. A *fuzzy classification system* $(\mathscr{C};\phi,\varphi,N)$ will be *meaningful* if and only if

- $\varphi\{\mu_c(x)/c\in\mathscr{C}\}>0,\ \forall x$,
- For every $\mathscr{A}\subset\mathscr{C}$ with at least two classes there exists $x\in X$ such that $\phi\{\mu_c(x)/c\in\mathscr{A}\}<1$.

Whenever $\varphi\{\mu_c(x)/c\in\mathscr{C}\}=0$ holds for some $x\in X$, then our classification system is claiming for a new class (at least one object cannot be included in any of the classes).

Whenever $\phi\{\mu_c(x),\mu_d(x)\}=1$ holds for all $x\in X$ and for some $c\neq d$, these two classes c and d are fully redundant (either c or d must be deleted).

As a general criteria, the higher value $\varphi\{\mu_c(x)/c\in\mathscr{C}\}$ and the lower values $\phi\{\mu_c(x)/c\in\mathscr{A}\}$ the better. The closer we are to such an extreme situation, taking the above indexes values 1 and 0 respectively, the closer we are to a *fuzzy partition*.

Definition 6.2. A *fuzzy partition system* is a *fuzzy classification system* $(\mathscr{C};\phi,\varphi,N)$ such that the associated membership functions verify the following two conditions:

1. $\varphi\{\mu_c(x)/c\in\mathscr{C}\}=1,\ \forall x$,
2. $\phi\{\mu_c(x)/c\in\mathscr{A}\}=0,\ \forall x$ and every $\mathscr{A}\subset\mathscr{C}$ containing at least two classes.

7. About the learning process

It is clear that a *nice* set of classes will in general be a result of a sometimes long learning process, with a sequence of arguments suggesting to delete some classes, to substitute classes by means of mixtures of previous classes, or to introduce extra classes. The above indices can give us useful hints about the behaviour of our classification system, and in turn lead to successive improvements. From this point of view, a *fuzzy partition* is just a fully explanatory and nonredundant *fuzzy classification system* that may be the final objective for many decision makers (but not necessarily a declared objective). Anyway, a natural learning processes will search for a better classification system (in a sense to be fixed by each user), taking into account at least the above *covering*, *relevancy* and *redundancy* arguments: try always to explain as much as you can, avoiding as much as possible irrelevant and redundant classes.

Of course each one of those three key arguments (covering, relevancy and redundancy) allow degrees of verification, and decision makers should make their mind up about the right levels they are willing to accept. For example, if there are objects $x\in X$ such that $\varphi\{\mu_c(x),c\in\mathscr{C}\}=0$, it suggests a search for some one extra class in order to capture information about those pixels. These pixels therefore deserve a more careful look in order to find out some missing class or some missing characteristic. Obviously, we shall be searching for new extra classes whenever we are close to this situation. Analogously, if $\varphi\{\mu_c(x),\mu_d(x)\}=1,\forall x\in X$, for two classes $c\neq d$, then we should be deleting one of those two classes. The closer to this situation, the stronger the need for a redefinition of the classes under consideration. In general, once covering has been checked, if

$$\varphi\{\mu_d(x)/d\neq c\}$$

is low, we should think if c as a relevant class. Analogously, if

$$\phi\{\mu_c(x),\mu_d(x)\}$$

is high, we should be thinking of *redefining* those classes c or d.

8. An application to a digital image classification

We shall consider here the same picture (127 pixels) analyzed by some of the authors in [3] (a LANDSAT 5 satellite image of southern Spain, *Worldwide Reference System* 202-34-4). Numerical analysis of the intensity bands of the three basic colors (red, green and blue) has been done with MATLAB packages.

First of all, we fixed our recursive triple (ϕ, φ, N). Taking into account the axiomatic justification of some particular recursive rules shown in [4], we consider a recursive triple such that

- $\phi_n(a_1, \ldots, a_n) = \dfrac{3 \cdot \Pi_{k=1}^{n} a_k}{1 + 2 \cdot \Pi_{k=1}^{n} a_k}$,
- $\varphi_n(a_1, \ldots, a_n) = \dfrac{1 - (\Pi_{k=1}^{n}(1 - a_k))}{1 + 2 \cdot \Pi_{k=1}^{n}(1 - a_k)}$,
- $N(x) = 1 - x$.

Notice that, in this case,

$$\varphi\{\mu_d(x)/d \neq c\} < 1 \quad \forall c$$
$$\Rightarrow \varphi\{\mu_c(x)/c \in \mathscr{A}\} < 1 \quad \forall A$$

and

$$\phi\{\mu_c(x), \mu_d(x)\} > 0 \quad \forall c \neq d$$
$$\Rightarrow \phi\{\mu_c(x)/c \in \mathscr{A}\} > 0 \quad \forall A.$$

The above recursive triple is a very particular case of the more general model studied in [4].

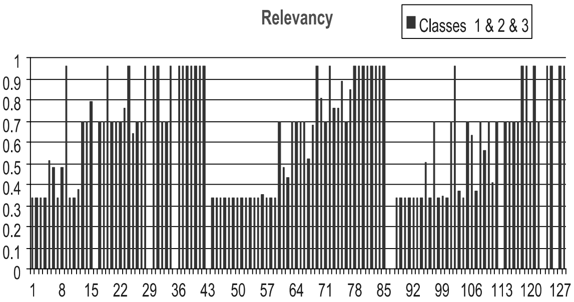
The image in [3] was analyzed applying the algorithm initially proposed in [2]. Three fuzzy classes were obtained. Centers for each class were automatically defined by means of a crisp k-means algorithm (see [5]), and fuzzy membership functions for each class were then obtained from those centers, following [2]. In this way a formal definition of a set of three classes, $\mathscr{C} = \{c_1, c_2, c_3\}$, was given in [3]. From this previous result we can now learn about the quality of such a classification, by considering some of the above indexes.

In particular, we show here the following graphics:

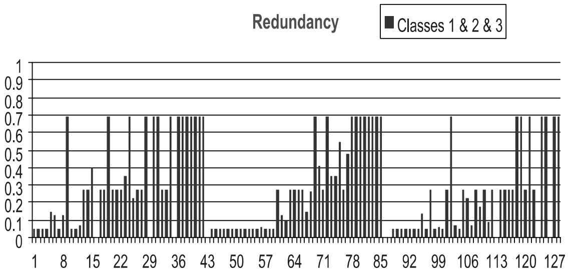
- Fig. 1(a) shows the pixel by pixel disjunctive aggregation of degrees of membership (relevancy of the three classes together, i.e., covering)

  $\varphi\{\mu_{c_1}(x), \mu_{c_2}(x), \mu_{c_3}(x)\}$.
- Fig. 1(b) shows the pixel by pixel conjunctive aggregation of degrees of membership (i.e., redundancy of the three classes together)

  $\phi\{\mu_{c_1}(x), \mu_{c_2}(x), \mu_{c_3}(x)\}$.



(a) Relevancy (all three classes)



(b) Redundancy (all three classes)

Fig. 1. Analysis for joint three classes.

- Fig. 2(a), (b) and (c) show the disjunctive aggregation (relevancy) values for each pair of classes, i.e.,

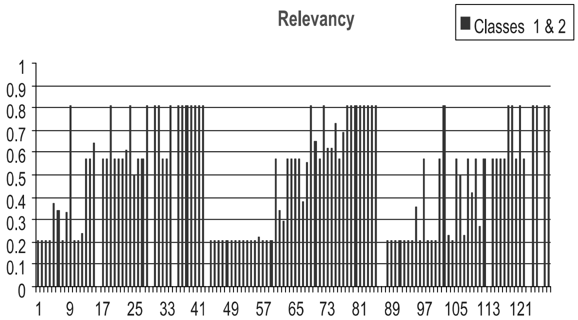
  $\varphi(\mu_{c_i}(x), \mu_{c_j}(x)) \quad \forall i \neq j$

  for each pixel x.
- Fig. 3(a), (b) and (c) show the conjunctive aggregation (redundancy) values for each pair of classes, i.e.,

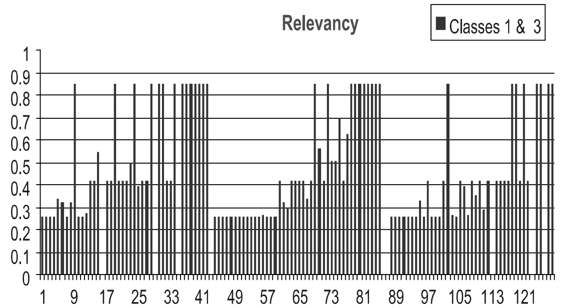
  $\phi(\mu_{c_i}(x), \mu_{c_j}(x)) \quad \forall i \neq j$

  for each pixel x.

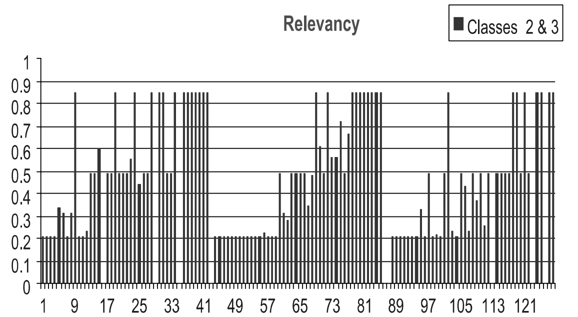
Notice that there are a number of pixels with low degree of explanation in terms of those three classes, even unclassified pixels (pixels with $\varphi\{\mu_{c_1}(x), \mu_{c_2}(x), \mu_{c_3}(x)\} = 0$). A deeper analysis of these pixels should allow hints about a possible new class or some missing characteristic. Some pixels seem to be quite well classified taking into account only two of those three classes (pixels with $\varphi\{\mu_{c_1}(x), \mu_{c_2}(x)\}$, $\varphi\{\mu_{c_1}(x), \mu_{c_3}(x)\}$ and $\varphi\{\mu_{c_2}(x), \mu_{c_3}(x)\}$ around 0.8 and 0.9); but some pixels (see, e.g., pixel 95) get a better classification in terms of



(a) Classes 1 and 2

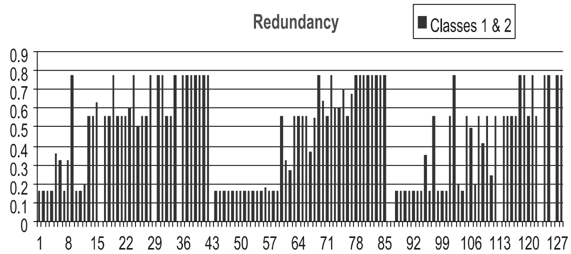


(b) Classes 1 and 3

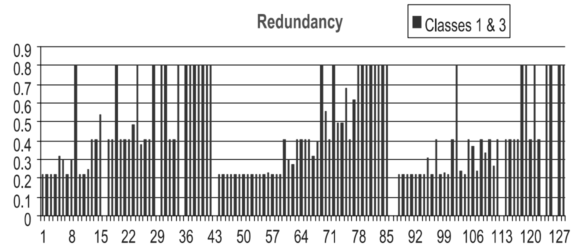


(c) Classes 2 and 3

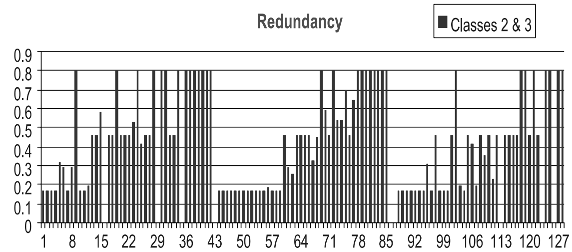
Fig. 2. Relevancy for pairs of classes.



(a) Classes 1 and 2



(b) Classes 1 and 3



(c) Classes 2 and 3

Fig. 3. Redundancy for pairs of classes.

classes 1 and 2 rather than classes 1 and 3 or 2 and 3. Moreover, high redundancy between classes appears for some pixels (even for pairs of classes), suggesting that classes can be redefined.

A careful look at each pixel, taking into account relevancy and redundancy behaviour, should give hints about how our family of classes can be improved (most probably with a manual search and not by means of any pure automatic procedure, including of course some statistical analysis). In practice, this stage needs an active participation of the decision maker, and this situation will be repeated each time we analyze a new family of classes. Nobody should expect a wonderful classification of a complex problem just in one iteration, but at least we have been able here to evaluate classification results according to several standard arguments.

9. Final comments

This paper describes some desirable properties linked to fuzzy classification systems. These properties are linked to a certain degrees of covering, relevancy and redundancy. In principle, relevancy and redundancy can be considered for any subset of classes, but in practice perhaps they can be applied to few cases, checking relevancy just for isolated classes and checking redundancy to every pair of classes, at least at a first stage (similar situation appears in probability, where independence of a family of events refers to every finite subfamily of those events). In fact, we do not expect this situation: depending on results we shall be considering additional indexes (again, statistical independence of random variables use to be approached in practice just by evaluating the matrix of linear correlation indexes).

The classification model proposed in this paper shows a promising behaviour. On one hand, covering and relevancy are analyzed by means of a disjunctive rule, leading to deletion of classes or a search for extra classes. On the other hand, redundancy is analyzed by means of a conjunctive rule, leading to a search for new mixtures of classes.

Together with the above indexes for covering, relevancy and redundancy, we can also take into account some global indexes in order to get an evaluation of the quality of a classification in those terms. For example, the value

$$\phi(\varphi\{\mu_c(x)/c \in \mathscr{C}\}, x \in X)$$

can be considered a measure of how explanatory our classification system is (about the set of objects X). Analogously, the value

$$\varphi(\phi\{\mu_c(x), \mu_d(x)\}, x \in X)$$

can be considered a measure of how redundant those two classes $c \neq d$ are (about the set of objects X). In general, we can fix certain levels of goodness for both global indexes, and repeat the whole process again and again till we get a good enough classification. Of course, decision maker should evaluate effort in searching for a new set of classes, and almost for sure there will be a maximum number of classes decision maker is willing to deal with.

Since in our example there are pixels showing zero degree of membership for all three classes under consideration, the above two global indexes will give a trivial information: on one hand,

$$\phi(\varphi\{\mu_{c_1}(x), \mu_{c_2}(x), \mu_{c_3}(x)\}, x \in X) = 0$$

and on the other hand,

$$\varphi(\phi\{\mu_{c_i}(x), \mu_{c_j}(x)\}, x \in X) = 0 \quad (i \neq j).$$

In this way, those three classes obtained in [3] should not be considered good enough: we should carefully analyze those unclassified pixels and search for an extra class in order to get some explanation for them (at a second stage we should have a look to those pixels showing high overlapping between classes).

Obviously, this example suggests that alternative aggregation operators should be tried. An appropriate OWA operator [34], for example, will allow to keep very low values for a fixed number of pixels (we should not pursue to explain *all* pixels, but only *most of them*, perhaps a certain percentage).

Much more research through real examples is indeed needed, in order to fix membership functions and aggregation rules. Any learning procedure in a complex context represents an extremely difficult issue. It will for sure require some *creative* abilities [6,27,30] (arguments discussed in [9] must be also taken into account). Of course, behaviour of surrounding pixels should also be taken into account. Moreover, as pointed out in [22,23], the operation rules within objects may be not the same as the operation rules within classes. Several disjunction operators may co-exist in our model in order to evaluate the set of classes every time it has been redefined.

Covering, relevancy and redundancy as introduced in this paper allow a quite general theoretical framework based upon aggregation rules. All previous already known results from aggregation functions can be taken into account, not only relative to t-conorms but to more general approaches based upon a unique associative binary operator [11–13,20,21,35]. On the contrary, our recursive rules needs not to be based upon a unique binary operator, but it may evolve *in time* (with consistent changes depending on the number of item being amalgamated till that moment). Notice the deep relation of our approach with the structure of fuzzy preferences, as developed in [24] (see also [15,16]).

Anyhow, final decision about our classification system and its possible improvements has to be made by at least considering covering, relevancy and redundancy, always keeping in mind that our main objective should be a better comprehension of the problem (sometimes we forget that complex information may be not manageable by the decision maker). In this sense, we do not expect that an optimal number of classes exist, because it will depend on each particular decision maker (although we can guess a general proposal after some experience in a particular context). As it is well known, imposing more homogeneous classes improves accuracy of the model, but at the same time the model becomes more complex (the number of classes use to increase).

Finally, it is important to point out again that a fuzzy partition (including classical Ruspini's definition) may be not a good classification system for some decision makers. In fact, some classification problems do require a family of classes with clear overlapping, even allowing objects fully belonging to several classes. As pointed out in [19], evaluation of the quality of a fuzzy partition is an intricate problem that still requires many developments. The approach proposed in this paper will hopefully help. Future applications will be developed in a remote sensing context, taking into account previous results obtained by some of the authors in [3].

Acknowledgements

This research has been partially funded by the Government of Spain, and by a joint University of California and Complutense University project within the Del Amo Program. We also acknowledge comments of an anonymous referee, who pointed out key issues which indeed improved the paper.

References

[1] A. Amo, J. Montero, V. Cutello, On the principles of fuzzy classification, In: Proceedings North American Fuzzy Information Processing Society Conference, 1999, pp. 675–679.

[2] A. Amo, J. Montero, G. Biging, Classifying pixels by means of fuzzy relations, International Journal of General Systems 29 (2000) 605–621.

[3] A. Amo, J. Montero, A. Fernández, M. López, J. Tordesillas, G. Biging, Spectral fuzzy classification: An application, IEEE Transactions Systems Man and Cybernautics (C) 32 (2002) 42–48.

[4] A. Amo, J. Montero, E. Molina, On the representation of recursive rules, European Journal of Operational Research 130 (2001) 29–53.

[5] G.H. Ball, D.J. Hall, ISODATA—A Novel Method of Data Analysis and Pattern Classification, Stanford Research Institute, Menlo Park, CA, 1965.

[6] V. Belton, J. Hodgkin, Facilitators, decision makers, D.I.Y. users: Is intelligent multicriteria decision support for all feasible or desirable, European Journal of Operational Research 113 (1999) 247–260.

[7] J.C. Bezdek, J.D. Harris, Fuzzy partitions and relations: An axiomatic basis for clustering, Fuzzy Sets and Systems 1 (1978) 111–127.

[8] D. Butnariu, Additive fuzzy measures and integrals, Journal of Mathematical Analysis and Applications 93 (1983) 436–452.

[9] V. Cutello, J. Montero, Nondeterministic fuzzy classification systems, In: Proceedings FUZZ-IEEE Conference, 1997, pp. 1689–1694.

[10] V. Cutello, J. Montero, Recursive connective rules, International of Journal of Intelligent Systems 14 (1999) 3–20.

[11] B. De Baets, Idempotent uninorms, European Journal of Operational Research 118 (1999) 631–642.

[12] J. Dombi, Basic concepts for a theory of evaluation: The aggregative operator, European Journal of Operational Research 10 (1982) 282–293.

[13] J. Dombi, A general class of fuzzy operators, Fuzzy Sets and Systems 8 (1982) 149–163.

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

[15] J. Fodor, M. Roubens, Valued preference structures, European Journal of Operational Research 79 (1994) 277–286.

[16] J. Fodor, M. Roubens, Fuzzy Preference Modelling and Multicriteria Decision Support, Kluwer, Dordrecht, 1994.

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

[18] I. Iancu, Connectives for fuzzy partitions, Fuzzy Sets and Systems 101 (1999) 509–512.

[19] P. Matsakis, S. Andrèfouët, P. Capolsini, Evaluation of fuzzy partitions, Remote Sensing of Environment 74 (2000) 516–533.

[20] R. Mesiar, Aggregation operators: Some classes and construction methods, In: Proceedings Eight International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2000, pp. 707–711.

[21] R. Mesiar, B. De Baets, New construction methods for aggregation operators, In: Proceedings Eight International Conference on Information Processing and Management

of Uncertainty in Knowledge-Based Systems, 2000, pp. 701–710.

[22] J. Montero, Comprehensive fuzziness, Fuzzy Sets and Systems 20 (1986) 86–89.

[23] J. Montero, Extensive fuzziness, Fuzzy Sets and Systems 21 (1986) 201–209.

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

[25] J. Montero, J. Yáñez, J. González-Pachón, Searching for the dimension of a fuzzy preference relation, EURO Conference Budapest, July 16–19, 2000.

[26] M. Roubens, Pattern classification problems and fuzzy sets, Fuzzy Sets and Systems 1 (1978) 239–253.

[27] B. Roy, Decision science or decision-aid science, European Journal of Operational Research 66 (1993) 184–203.

[28] E.H. Ruspini, A new approach to clustering, Information and Control 15 (1969) 22–32.

[29] B. Schwizer, A. Sklar, Probabilistic Metric Spaces, North-Holland, New York, 1983.

[30] G. Shafer, Savage revisited, Statistical Science 1 (1990) 463–501.

[31] H. Thiele, A characterization of Ruspini-partitions by similarity relations, In: Proceedings International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 1996, pp. 389–394.

[32] H. Thiele, A characterization of arbitrary Ruspini-partitions by fuzzy similarity relations, In: Proceedings FUZZ-IEEE Conference, 1997, pp. 131–134.

[33] E. Trillas, On negation functions in fuzzy set theory, in: S. Barro, A. Bugarín, A. Sobrino (Eds.), Advances in Fuzzy logic, Universidade de Santiago de Compostela, 1998, pp. 31–43.

[34] R.R. Yager, On ordered weighted averaging aggregation operators in multi-criteria decision making, IEEE Transactions on Systems, Man and Cybernetics 18 (1988) 183–190.

[35] R. Yager, A. Rybalov, Uninorm aggregation operators, Fuzzy Sets and Systems 80 (1996) 111–120.

[36] J. Yáñez, J. Montero, A poset dimension algorihtm, Journal of Algorithms 30 (1999) 185–208.

[37] L.A. Zadeh, Fuzzy sets, Information and Control 8 (1965) 338–353.

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