ON THE PRINCIPLES OF FUZZY CLASSIFICATION

A. del Amo, J. Montero
Faculty of Mathematics
Complutense University
Madrid, Spain

V. Cutello
Dept. of Mathematics
University of Catania
Catania, Italy

Abstract

In this paper we analyze the concept of fuzzy partition, *starting from the classical key definition given by Ruspini. Our main claim is that such a definition is too restrictive, since it assumes a particular set of classes that in practice may be reached only after a long learning process. As a consequence, some principles to be taken into account in fuzzy classification methods are discussed.*

1. Introduction.

Classification and Control have been since the beginning of Fuzzy Set Theory [17] two central fields for their theoretical and practical developments (see, e.g., [3, 18]). In fact, many problems within both fields are very naturally formalized by introducing fuzzy concepts. In some cases, a fuzzy approach seems to offer a useful simplification of a too complex reality. As it is the case for control problems. In other cases, the concepts that users have in mind are fuzzy in nature. 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 of direct observations.

In many fuzzy classification applications, a set of classes $\mathcal{C}$ is assumed. The problem is then to determine for every object x under consideration, $x \in X$, the degree $\mu_c(x)$ to which object x belongs to class $c \in \mathcal{C}$. In this way, a membership function $\mu_c : X \to [0,1]$ has been defined for each class $c \in \mathcal{C}$ (see, e.g., Roubens [11]).

Such an approach seems to us still unrealistic, since most users will find serious difficulties in assigning degrees of membership to one class without taking into consideration the remaining possibilities for classification. Classification methods are in practice highly dependent on the closed family of classes the user is forced to consider (even in a crisp context, users frequently look at all possible choices before choosing a particular class for a given object).

A key concept in classification is therefore the notion of partition, since it produces a structured family of classes (classes are strongly related between them).

Fuzzy partitions were introduced by Ruspini [12] (see also Bezdek-Harris [4]): given a discrete family $\mathcal{C}$ of *classes*, it is assumed that for every object $x \in X$ under consideration $\sum_{c \in \mathcal{C}} \mu_c(x) = 1$ always holds. Each object may belong to several classes -to a certain degree- 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, then each object will be in one and only one class:

Ruspini's proposal represents from our point of view a desirable situation in many situations, but it appears to be again a too restrictive definition within fuzzy modelling. In most cases, the fuzzy classes under consideration do not produce such a Ruspini fuzzy partition, and it is only perhaps after a long learning process that users are able to get a new family of fuzzy classes, assuring that every object is fully explained with no superflous information.

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, due to market restrictions use to require a large number of classes, and each piece of fruit is allowed to simultaneously be associated to several different classes.

Some practical difficulties of Ruspini's partitions can

0-7803-5211-4/99/$10.00

be partially overcome by a weaker approach proposed by some other authors (see, e.g., [13, 14]). In this paper we propose to analize classification systems by means of aggregative models, which should present Ruspini's partition as a particular additive solution. Then we can take into account some results already obtained in other formal contexts in order to axiomatically explain different classification structures, each one justified by means of a particular aggregation operator (see, e.g., [6, 7, 10] but also [16]).

2. Fuzzy Classification system.

Let us assume a fixed finite set of objects X. A ***fuzzy classification system*** will mean here a finite family $\mathcal{C}$ of fuzzy classes, each $c \in \mathcal{C}$ with its associated membership function $\mu_c : X \to [0,1]$ together with a De Morgan triple (T, S, n), where

- $S : [0,1] \times [0,1] \to [0,1]$ is a t-conorm, i.e., an associative and commutative nondecreasing continuous mapping such that $S(0,1) = 1$, playing the role of a disjunction function (see [6, 7, 10]),
- $n : [0,1] \to [0,1]$ is a negation function, i.e., a strictly decreasing associative continuous function such that $n(0) = 1$ and $n(1) = 0$ (see [15]),and
- $T : [0,1] \times [0,1] \to [0,1]$ is the t-norm defined as

  $$T(x,y) = n(S(x,y) \qquad \forall x,y$$

  in such a way that it plays the role of a conjunction function (a commutative nondecreasing associative continuous mapping such that $T(0,1) = 0$).

Then, the degree to which an object $x \in X$ belongs to such a family of classes $\mathcal{C}$ will be, depending on the particular aggregation operator S being previously chosen,

$$S\{\mu_c(x)/c \bullet \mathcal{C}\}$$

and such a value can be understood as the degree to which such an object is *explained* by such a family of classes $\mathcal{C}$. The higher all these values $\mu_{\mathcal{C}}(x)$ are, the better. The whole famlily of objects should be in this way *covered* by the set of fuzzy classes.

But it is obvious that such an numerical analysis may lead to trivial paradoxes, for example by replicating several times a unique class.

At any rate, the family of classes should be as simple (compact) as possible, taking into account only those ***relevant*** classes, but also reducing all possible overlapping information (***redundancy***). Let us discuss these two notions in the next sections.

3. Relevance

Of course a ***void*** class c such that $\mu_c(x) = 0 \quad \forall x \in X$ should be deleted from the model.

The same happens when classes are replicated, i.e., when $\mu_c(x) = \mu_k(x)$ $\forall x \in X$ being c, k two different classes ($c \neq k$).

The above two ***crisp*** situations can be easily solved, 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 (crisp) cases. In both two cases, a class is ***almost*** useless just because it barely helps to explain anything (at least when faced to our particular fixed set of objects).

In Thiele [13, 14], for example, void classes are excluded axiomatically. Certainly, we should take into account in our model only those classes c being ***relevant*** in the sense that $\mu_c(x) > 0$ for some object x. But being ***relevant*** is a matter of degree.

A first proposal in order to measure ***relevance*** of class $c \in \mathcal{C}$ is to evaluate $S\{\mu_c(x)/x \in X\}$ and we can think that a class c is in principle less important as far as such a value is lower. But such an approach is misleading:

- $\mu_c(x)$ may be low, but still the only thing we know about object x : that 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 giving us any information by itself: for example, if $\mu_c(x) = \mu_c(y), \forall y \neq x$.
- $\mu_c(x)$ may be high, but still that object x is much better described by other classes: for example, if $\mu_k(x) > \mu_c(x)$ and $\mu_k(y) = \mu_c(y), \forall y \neq x$.

Relevance should always be evaluated not for each isolated class c ***to be deleted***, but for the selected set of classes $\mathcal{C} - \{c\}$ ***to be kept*** in the model, by comparing

$$S\{\mu_c(x)/c \in \mathcal{C}\} \text{ with } S\{\mu_k(x)/k \in \mathcal{C}, k \neq c\}$$

for every object x. This approach represents the basis for a formal analysis of ***relevance***, searching for the smallest set of classes which maintains the user's desired explanatory properties.

The ***relevance*** issue can be therefore addressed as a ***dimensionality reduction*** problem. 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 still keeping enough explanatory power. In practice we shall always look for an *appropriate* number of classes (the lower number of classes, the better) explaining *most* of the problem we are faced to.

Notice that the user may be willing to accept that objects belong to several classes, even a clear crisp overlaping ($\mu_c(x) = \mu_k(x) = 1$ for some $c \neq k$), if such a situation is considered as *relevant* for the real decision making problem classification the user will be faced with. 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.

4. Redundancy

Once the initial set of classes has been analyzed and *non-relevant* classes have been suppresed from the model, we can assume that every class is giving us some kind 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 *representation system* of the set of objects. *Redundancy* suggests the possible existence of an alternative representation, to be found by re-defining our classes. It must be taken into account how difficult is to re-define classes or define extra ones. In general, classes need to have some real meaning for decision makers: otherwise they will not be able to assign degrees of membership.

Once *relevancy* has been studied at a first stage, class overlapping will be searched by means of the above t-norm T (see [10]). In fact, the value

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

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

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

1. $S\{\mu_c(x)/c \in \mathcal{C}\} = \max\{\mu_c(x)/c \in \mathcal{C}\} = 1, \forall x$, and
2. $T\{\mu_c(x), \mu_k(x)\} = \min\{\mu_c(x), \mu_k(x)\} = 0, \forall x \in X, \forall c \neq k$

(we know the only available crisp t-conorm is the *max* operator).

Analogously, when dealing with Ruspini partitions,

$$\mu_c(x) = a \Rightarrow \sum_{k \neq c} \mu_k(x) = 1 - a$$

and, by choosing Luckasievich's t-conorm as the aggregation function S, we have

1. $S\{\mu_c(x)\} = \min\{\sum_{c \in \mathcal{C}} \mu_c(x), 1\} = 1, \qquad \forall x$, and
2. $T\{\mu_c(x), \mu_k(x)\} = \max\{0, \mu_c(x) + \mu_k(x) - 1\} = 0, \qquad \forall x, \forall c \neq k$

A *fuzzy classification system* can be therefore denoted by

$$(\mathcal{C}; S, T, n)$$

Covering and overlapping as introduced in this paper allow a theoretical framework based upon aggregation rules. All previous already known results from aggregation functions can be taken into account, not only restricted to t-conorms but more general approaches ([6, 7, 10, 5, 16]).

A more general model for to fuzzy classification systems, which avoids associativity as a basic theoretical hypothesis, would require for the basic assumptions an approach closer to the preference model developed in [10]. Such a more general approach will be addressed in the future.

5. Fuzzy partition systems

A *fuzzy classication system* $(\mathcal{C}; S, T, n)$ will be obviously 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.

DEFINITION 5.1
A fuzzy classication system $(\mathcal{C}; S, T, n)$ *will be meaningful if and only if*

1. $S\{\mu_c(x)/c \in \mathcal{C}\} > 0, \forall x$, *and*
2. $T\{\mu_c(x), \mu_k(x)\} < 1, \forall x \in X, \forall c \neq k$ ■

Whenever

$$S\{\mu_c(x)/c \in \mathcal{C}\} = 0$$

for some x, or

$$T\{\mu_c(x), \mu_k(x)\} = 1, \forall x \in X$$

for some $c \neq k$, it means, respectively, that object x is out of our classification system (at least one extra class is needed), or those classes c, k are fully redundant (at least one class can be deleted).

As a general criterion, the higher all values

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

and the lower all values

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

the better. The closer we are to this extreme situation, taking the above indexes values 1 and 0 respectively, the closer we are to the true notion of *fuzzy partition*.

DEFINITION 5.2 *A* fuzzy partition system *is a* fuzzy classication system $(\mathcal{C}; S, T, n)$, *such that the associated membership functions verify the following two conditions:*

1. $S\{\mu_c(x)/c \in \mathcal{C}\} = 1, \forall x$, *and*

2. $T\{\mu_c(x), \mu_k(x)\}) = 0, \forall c \neq k, \forall x$ ■

Obviously, the values

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

become, respectively, a measure of how explanatory our classification system is and how redundant our classification system is (about object x).

6. About the learning process

It is clear that a *nice* set of classes will only be a result of a sometimes long learning process, deleting, modifying or introducing new classes.

The above indices give us some key hints for the necessary learning process about our classes behaviour, which in turn should lead to a better *classification system*. A *fuzzy partition system* is just a fully explanatory and non-redundant *fuzzy classification system*.

Natural learning processes that search for a better classification system suggest the above *covering*, *relevance* and *redundancy* arguments: try always to explain as much as you can, not taking into account the most irrelevant and redundant classes.

Of course each one of those three key arguments (covering, relevance and redundancy) allows degrees of verification, and decision makers should make up their mind about the right levels they are willing to accept. Let us have a look at some other basic properties we all are willing to impose to our classification model. Subsequent *fuzzy* properties should hold the closer we are to such a situation.

1. If a class c is void, i.e., $\mu_c(x) = 0, \forall x \in X$, then the explanatory level must not change if such a class is deleted. If a class is close to this situation, the explanatory level should be accordingly modified when the class is suppressed.
2. If two classes c, k are such that $\mu_c(x) = \mu_k(x), \forall x \in X$, then the explanatory level must not change if one of those classes is deleted. If two classes are close to this situation, the explanatory level should accordingly be modified when one of them is suppressed.
3. If there is an object x such that $\mu_c(x) = 0, \forall c \in \mathcal{C}$ (i.e., $S\{\mu_c(x)/c \in \mathcal{C}\} = 0$), then there is an absolute need for at least one extra class. If we are close to this situation, we should start a search for new extra classes.
4. If $T\{\mu_c(x), \mu_k(x)\} = 1$ for some object, then there is some overlaping and depending on the other objects we may think of redefining the set of classes. The closer to this situation the stronger the need for a redefinition of the classes under consideration.

In general, once a set of poorly informative classes have been eliminated, if

$$S\{\mu_c(x)/c \in \mathcal{C}\} < 1$$

we should think of introducing *extra* classes. Analogously, if

$$T\{\mu_c(x), \mu_k(x)\} > 0$$

we should be thinking of *redefining* previous classes.

Learning procedures represent indeed a complex problem, since they require not only the above useful hints but some *creative* abilities.

Another key point which will be addressed in the future is the dependency of the obtained fuzzy classification system on the history of the learning process. For instance, deleting first one class c instead of c' and readjusting the explanatory level, may force us to delete other classes or to introduce new classes. Will we obtain the same result when we start by deleting first the class c' ?

7. Final comments

The approach proposed in this paper is actually being tested in a particular Remote Sensing classification problems (see [1] for a first approach), taking into account the above three different characteristics: covering (leading to the deletion of some classes), relevancy (leading to the introduction of some extra classes), and redundacy (leading to a re-definition of classes).

It is important to realize that a Ruspini partition may be not a good classification system for the decision maker. Some classification problems do require a family of classes with clear overlaping, even allowing objects fully belonging to several classes ($\mu_c(x) = \mu_k(x) = 1$ for some $c \neq k$). Global indexes for covering, relevancy and overlaping can be obtained by direct aggregation along objects in X.

Future research should allow an axiomatic justification of the model not directly assuming the existence of a De Morgan triple, but a more meaningful set of assumptions under this particular context (even considering aggregation rules not based upon t-conorms and that need some extra mathematical formalization in order to ensure some sort of associativity property [5]). Moreover, as pointed out in [8, 9], the operation rules within objects may be not the same as the operation rules within classes, as we have imposed here. Several disjunction operators may co-exist in the model.

Acknowledgements: This research has been partially supported by Dirección General de Investigación Científica y Técnica (Spain), project PB95-0407, and also by the NATO collaborative research grant CRG-960242.

References

[1] A. del Amo, J. Montero and G. Biging, Classifying pixels by means of fuzzy relations, *Int. J. General Systems* (to appear)

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

[3] J.C. Bezdek, 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, 1992); pages 1–27

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

[5] V. Cutello and J. Montero, Recursive connective rules, *Int. J. Intelligent Systems* vol. 14(1), pg. 3–20, 1999

[6] J. Dombi. 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 (1982).

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

[8] J. Montero, Comprehensive fuzziness, *Fuzzy Sets and Systems* 20:79–86 (1986)

[9] J. Montero, Extensive fuzziness, *Fuzzy Sets and Systems* 21:201–209 (1987)

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

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

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

[13] H. Thiele, A characterization of Ruspini-partitions by similarity relations, *Proceedings I.P.M.U.'96*, pages 389–394 (1996).

[14] H. Thiele, A characterization of arbitrary Ruspini-partitions by fuzzy similarity relations, *Proccedings FUZZ-IEEE*, pages 131–134 (1997).

[15] E. Trillas, Sobre funciones de negacion en la teoria de conjuntos difusos. *Stochastica*, 3:47–59 (1979).

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

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

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