STUDIA UNIV. BABEŞ-BOLYAI, MATHEMATICA, XL, 3, 1995

FUZZY DECISION SUPERVIZED CLASSIFIERS

Horia F. POP*

Received: February 1, 1995
AMS subject classification: 68P99

Rezumat. Clasificatori supervizaţi cu decizie nuanţată. In acest articol sunt prezentate o serie de generalizări ale unor algoritmi de instruire clasici. Sunt prezentate câteva din problemele generate de aceştia. Apoi se construieşte un algoritm de clasificare supervizată nuanţată bazat pe o generalizare a algoritmului Fuzzy n-Medii. Sunt studiate proprietăţile sale şi sunt trecute în revistă câteva avantajele obţinute prin utilizarea sa.

1. Introduction

Let us consider a set of objects, $X = \{x^1, ..., x^p\} \subset \mathbf{R}^d$, classified with a fuzzy clustering algorithm of the type Fuzzy n-Means, and the fuzzy partition $P = \{A_1, ..., A_n\}$ coresponding to the cluster substructure of the set X (see [2,3]).

We rise the problem of including an extra-object $x^0 \notin X$ in the cluster structure of X. Of course, this would mean to determine the membership degrees of x^0 to the fuzzy sets members of the partition P. These degrees will provide sufficient information in order to classify the object x^0 with respect to the elements of X.

The algorithms that solve this kind of problems we be called *fuzzy decision supervized classification algorithms*. Supervised classification because the classification of the extra-object is realized using not only the data set X, but also the fuzzy partition obtained by classifying the set X. Fuzzy decision because, unlike the traditional classifiers, where the aim is to state in which classical subset the object may be included, now we are interested in the membership degrees of the object to the fuzzy sets members in the given fuzzy partition.

* *"Babeş-Bolyai" University, Faculty of Mathematics and Computer Science, 3400 Cluj-Napoca, Romania*

The simplest approach is the classification of the extended set $X \cup \{x^0\}$ using one of the common fuzzy clustering algorithms (see, for instance, [3]), and the comparation of the produced partition to the partition P. Although, this method is very costly considering the neccessary execution time, because it supposes the classification of the objects of X, set that in the real applications may be quite large. Also, this is not supervised classification, because the information provided by the fuzzy partition P is not used.

The alternative approach is to keep unchanged the membership degrees of the objects in X to the sets of the fuzzy partiton P, and to determine the membership degreess of x^0 as a consequence of the minimization of an objective function similar to those used for the algorithms of the type Fuzzy n-Means.

Also, we will present in this paper some more straightforward algorithms of this type. These algorithms are fuzzy generalizations of the well-known **k nearest neighbours and nearest prototype**.

2. The algorithm of the k nearest neighbours

The algorithm of the k nearest neighbours is one of the standard methods of the supervised classification and it has been especially remarked because of its simplicity. The method is based on the evaluation of the memberships of an unknown object to the classes of the given set using the distances between this object and its k nearest neighbours, and the memberships of these neighbours. The method presented here is a development of the simpler variant that attributes an object to the class that contains its nearest neighbour.

Let us consider a set of classified objects, $X = \{x^1, ..., x^p\} \subset \mathbf{R}^d$ and the fuzzy partition $P = \{A_1, ..., A_n\}$ corresponding to the cluster substructure of the set X. Let us consider $x^0 \in \mathbf{R}^d$ an object that needs to be classified with respect to the fuzzy partition P.
The classification of the object x^0 will be realized by examining a number of k objects from X which are the nearest ones to x^0 (the most similar).

Even if by this method will be produced n new fuzzy sets on $X \cup \{x^0\}$, denoted $\tilde{A}_i$ and having the property $\tilde{A}_i(x^j) = A_i(x^j)$ for every $i=1,...,n$ and $j=1,...,p$, the new fuzzy sets will also be denoted by A_i.

In what follows we will consider a dissimilarity measure on $X \cup \{x^0\}$ (see [3,5]). Let d be a metric in the $\mathbf{R}^d$ space.

The *dissimilarity* between the object x^0 and a certain object x^j of the set X may be defined as

$$D(x^0, x^j) = d^2(x^0, x^j), \; j=1,...,p \tag{1}$$

and may be interpreted as a measure of the "non-ressemblence" between the objects x^0 and x^j.

Let σ be a permutation of the set $\{1,...,p\}$, so that

$$D(x^0, x^{\sigma(i)}) \leq D(x^0, x^{\sigma(j)}) \leftrightarrow i \leq j. \tag{2}$$

In order to determine the membership degrees of the object x^0 we will need to take into consideration with differnt weights the membership degrees of the k nearest neighbours ($x^{\sigma(j)}$, $j \leq k$). The more similar $x^{\sigma(j)}$ is to x^0, the greater weight its memberships should be given. This remark leads us to the following empirical rule for computing the memberships of x^0:

$$A_i(x^0) = \frac{\sum_{j=1}^{k} \frac{A_i(x^{\sigma(j)})}{D(x^0, x^{\sigma(j)})}}{\sum_{j=1}^{k} \frac{1}{D(x^0, x^{\sigma(j)})}} \tag{3}$$

The obtained algorithm is called **the algorithm of the k nearest neighbours**:

S1 Be given X, P, k.

S2 Computes the dissimilarities with respect to the relation (1).

S3 Determines the permutation σ such that the condition (2) be verified.

S4 Computes the membership degrees of the object x^0 with respect to the relation (3).

Remark. The membership degrees computed in this way verify the relation

$$\sum_{i=1}^{n} A_i(x^0)=1.$$

Remark. An alternative to this method is not to take into account all the objects in X, but only a subset of them, namely the most representative ones, eventually those having the membership degree to one of the classes at least 0.8.

3. The algorithm of the nearest prototype

Let us consider a set of classified objects, $X = \{x^1,...,x^p\} \subset \mathbf{R}^d$ and the fuzzy partition $P = \{A_1,..., A_n\}$ corresponding to the cluster substructure of the set X. Let $x^0 \in \mathbf{R}^d$ be an object that needs to be classified with respect to the fuzzy partition P.

In what follows we will suppose that the clusters of the set X have a hyperspherical shape. Moreover, we will suppose that in order to identify the optimal fuzzy partition P, the Fuzzy n-Means algorithm has been used. We consider the clusters as being represented by puctual prototypes. We will denote the prototype of the class A_i with L^i, $L^i \in \mathbf{R}^d$. As stated by the Fuzzy n-Means algorithm, the expression of the prototypes L^i is given by

$$L^i=\frac{\sum_{j=1}^{p}(A_i(x^j))^2x^j}{\sum_{j=1}^{p}(A_i(x^j))^2}. \tag{4}$$

The unknown object x^0 will be classified with respect to the distance between it and the prototypes L^i of the classes members of the partition P. We will consider that the membership degree of the object x^0 to a certain class is as larger as the object is nearer the prototype of that class.

In this case we will also use the same notation A_i to denote the extended fuzzy sets, defined over $X \cup \{x^0\}$.

In what follows we will consider a dissimilarity measure over $X \cup \{x^0\}$. Let d be a metric in the $\mathbf{R}^d$ space.

The *dissimilarity* between the object x^0 and the prototype L^i of the set A_i is defined as

$$D_i(x^0, L^i) = (d_i(x^0, L^i))^2,$$

where d_i is the local metric induced by the metric d and by the fuzzy set A_i. The dissimilarity D may be interpreted as a measure of "nonresemblance" between the object x^0 and the prototype L^i. Using the definition (see [5]), it may be written as:

$$D_i\}(x^0, L^i) = (A_i(x^0))^2 \, d^2(x^0, L^i). \tag{5}$$

The inadequacy between the memberships $A_i(x^0)$ of the object x^0 to the n classes and the prototypes L^i of these classes may be written using the function $J(A_1(x^0),...,A_n(x^0))$ given by

$$\begin{aligned} J &= \sum_{i=1}^{n} D_i(x^0, L^i) \\ &= \sum_{i=1}^{n} (A_i(x^0))^2 d^2(x^0, L^i). \end{aligned} \tag{6}$$

So our problem may be reduced to the determination of those membership degrees $A_i(x^0)$ which minimize the objective function J. This result is stated by the following

Theorem. The membership degrees $A_i(x^0)$, $i=1,...,n$, are a minimum of the function J if and only if

$$A_i(x^0)=\frac{1}{\sum_{k=1}^{n}\frac{d^2(x^0,L^i)}{d^2(x^0,L^k)}} \tag{7}$$

The proof of this theorem, as the proofs of all the theorems in this paper, are very similar to the proofs of the minimality theorems presented in [2,3]. For this reason we will not give here any explicit proof.

The algorithm obtained using this theorem will be called **the algorithm of the nearest prototype**:

S1 Be given X and P.

S2 Determine the prototypes L^i using the relation (4).

S3 Computes the distances $d(x^0, L^i)$ from the object x^0 to the prototype L^i of the class A_i.

S4 Computes the membership degrees of the object x^0 with respect to the relation (7).

Remark. The membership degrees $A_i(x^0)$ computed using the relation (7) verify the initial supposition: the smaller the distance from x^0 to a prptptype is, the greater the membership degree of x^0 to that class will be.

Remark. The computational effort is more reduced at this method as compared to the previous one, if we take into account the fact that generally the number of classes is very much smaller than the number of object: it is enough to compute n distances as compared to p distances for the previous case.

Remark. The geometrical locus of the points x^0 characterized by equal memberships to the classes A_i and A_j, $i,j=1,\dots,n$, $i\neq j$, is the median hyperplane of the segment L^iL^j (which contains the points equally distanced from L^i and L^j). As a consequence, this algorithm presents the problem of the inequal clusters, considering that a point x0 from the outer part of a greater class has many chances to be captured by a smaller neighbour class.

Remark. In order to overpass the problem of inequal clusters, when producing the fuzzy partition P the adaptive version of the Fuzzy n-Means algorithm may be used (see [2]). The local adaptive distance defined for that algorithm may also be used to determine the membership degrees $A_i(x^0)$.

4. The Restricted Fuzzy n-Means algorithm

Let us consider a set of classified objects, $X=\{x^1,...,x^p\} \subset \mathbf{R}^d$ and the fuzzy partition $P=\{A_1,...,A_n\}$ corresponding to the cluster substructure of the saet X. Let $x^0 \in \mathbf{R}^d$ be an object that needs to be classified with respect to the fuzzy partition P.

Let us suppose that the partition P has been produced using the Fuzzy n-Means algorithm. Our aim is to develop an algorithm that should compute the optimal fuzzy partition $\bar{P}$ corresponding to the set $\bar{X} = X \cup \{x^0\}$, by using a mechanism of the type of Fuzzy n-Means, with the difference that the membership degrees of the objects in X to the classes A_i, $i=1,...,n$ may not be modified.

In what follows we will consider a metric d in the Euclidean space $\mathbf{R}^d$. We will suppose that d is norm induced, so

$d(x, y) = (x-y)^T M (x-y)$, $\forall$ x, y $\in \mathbf{R}^d$, where M is a symmetrical and positively defined matrix.

Let us remember that the objective function of the Fuzzy n-Means algorithm is

$$J(P,L)=\sum_{i=1}^{n}\sum_{j=1}^{p}(A_i(x^j))^2 d^2(x^j,L^i), \tag{8}$$

where L^i is the pointly prototype associated to the fuzzy class A_i.

So, the objective function we will have in mind in this case is

$$\bar{J}(\bar{P},L)=\sum_{i=1}^{n}\sum_{j=0}^{p}(A_i(x^j))^2 d^2(x^j,L^i), \tag{9}$$

with the mention that $A_i(x^j)$ are kept constant for each i and for $j=1,...,p$.

The classification problem reduces to the determination of the fuzzy partition $\tilde{P}$ and of the representation L that minimizes the function $\tilde{J}$. The main result with this respect is given by the following

Theorem. (i) The fuzzy partition $\tilde{P}=\{A_1,...,A_n\}$ is minimum of the function $\tilde{J}(., L)$ if and only if

$$A_i(x^0)=\frac{1}{\sum_{k=1}^{n}\frac{d^2(x^0,L^i)}{d^2(x^0,L^k)}} \tag{10}$$

(ii) The set of prototypes $L=\{L^1,...,L^n\}$ is minimum of the function $\tilde{J}(\tilde{P}, .)$ if and only if

$$L^i=\frac{\sum_{j=0}^{p}(A_i(x^j))^2x^j}{\sum_{j=0}^{p}(A_i(x^j))^2}. \tag{11}$$

With this result, the optimal membership degrees of x^0 to the classes A_i will be determined using an iterative method in which $\tilde{J}$ is succesively minimized with respect to $\tilde{P}$ and L. The process will start with the initialization of the prototypes L^i to the values corresponding to the optimal positions computed for the function J. The resulted algorithm, the **Restrictive Fuzzy n-Means Algorithm**, is the following:

S1 Be given X and P.

S2 Determine the initial positions of the prototypes L^i as the optimal positions computed for the function J.

S3 Determine the membership degrees $A_i(x^0)$, $i=1,...,n$, using the relation (10).

S4 Determine the new positions of the prototypes L^i using the relation (11).

S5 If the new prototypes are closed enough to the former ones, then stop, else go back to the step S3.

In what follows we propose to determine the geometrical locus of the points x^0 for which the membership degrees to two classes A_{i1} and A_{i2} are equal. Let us denote

$$A_{i1}(x^0) = A_{i2}(x^0) = a. \tag{12}$$

Firstly, let us denote by L^{*i}, $i=1,...,n$ the protptypes corresponding to the fuzzy partition P over X, as they have been computed using the Fuzzy n-Means Algorithm. Thus,

$$L^{*i} = \frac{\sum_{j=1}^{p} (A_i(x^j))^2 x^j}{\sum_{j=1}^{p} (A_i(x^j))^2}. \tag{13}$$

Let us denote by α_i^2 the following value:

$$\alpha_i^2 = \sum_{j=1}^{p} (A_i(x^j))^2.$$

Then, between the prototypes L^i and L^{*i} exists the relationship

$$x^0 - L^i = (x^0 - L^{*i}) \cdot \frac{\alpha_i^2}{\alpha_i^2 + (A_i(x0))^2}. \tag{14}$$

The relationship above may be written under the form

$$d(x^0, L^i) = d(x^0, L^{*i}) \cdot \frac{\alpha_i^2}{\alpha_i^2 + (A_i(x0))^2}, \tag{15}$$

or

$$d(x^0, L^{*i}) = d(x^0, L^i) \left(1 + \frac{(A_i(x^0))^2}{\alpha_i^2} \right). \tag{16}$$

The relation (10) which computes the membership degrees may be rewritten as follows:

$$A_i(x^0) = \frac{\frac{1}{d^2(x^0, L^i)}}{\sum_{k=1}^{n} \frac{1}{d^2(x^0, L^k)}} \tag{17}$$

(ii) The set of prototypes L={

From the relations (12) and (17) it results that

$$d(x^0, L^{i1}) = d(x^0, L^{i2}).$$

By relating to (16), it results

$$\frac{d(x^0, L^{*i_1})}{d(x^0, L^{*i_2})} = \frac{1+\frac{a}{\alpha_{i_1}^2}}{1+\frac{a}{\alpha_{i_1}^2}} = K, \tag{18}$$

where

$$K \begin{cases} <1 & if\ \alpha_{i_1}^2 > \alpha_{i_2}^2 \\ =1 & if\ \alpha_{i_1}^2 = \alpha_{i_2}^2 \\ >1 & if\ \alpha_{i_1}^2 < \alpha_{i_2}^2 \end{cases} \tag{19}$$

So, we have obtained the following

Theorem. The geometrical locus of the points x^0 having equal memberships to the classes A_{i1} and A_{i2} is on a hypersphere with the center in the point

$$c = L^{*i_2}\frac{K^2}{K^2-1} - L^{*i_1}\frac{1}{K^2-1}$$

and with the radius

$$r = d(L^{*i_1}, L^{*i_2})\frac{K}{|K^2-1|},$$

where K is that stated in the relation (19).

At a more careful analysis of this hypersphere's equation, we may do some interesting remarks.

Remark. This hypersphere "catches" inside itself the prototype

L^{*i} of the class having the largest index α_i^2.

Remark. This may be the main problem of the method and it indicates a dependency with respect to the dimensions of the classes. Although the effect seems to be even more important than to the traditional Fuzzy n-Means algorithm, the experiments done with this algorithm on test examples show that the effects have simmilar dimensions.

Remark. The problem may be solutioned in the same way the problem of the inequal sized clusters (see [2]) has been solved for the Fuzzy n-Means algorithm, namely by using an adaptive metric, with respect to which all the clusters having equal dimensions.

Remark. The center of the hypersphere is always outside the segment $L^{*i1}\,L^{*i2}$.

Remark. The hypersphere intersects the segment $L^{*i1}\,L^{*i2}$ in the point

$$v = L^{*i_1}\frac{1}{K+1} + L^{*i_2}\frac{K}{K+1}$$

which is the more distanced by the center of the segment the more distanced by 1 is K.

Remark. The hypersphere is entirely situated on one side of the median hyperplane of the segment $L^{*i1}\,L^{*i2}$.

Remark. Moreover, for K = 1, the hypersphere degenerates to the median hyperplane specified above.

REFERENCES

1. James C. Bezdek. Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.

2. D. Dumitrescu. Hierarchical pattern classification. Fuzzy Sets and Systems 28 (1988), 145--162.

3. D. Dumitrescu. Classification Theory. Univ. of Cluj-Napoca Press, 1991.

D. Dumitrescu, Horia F. Pop. Convex decomposition of fuzzy partitions, I,II. Fuzzy Sets and Systems (1995). to appear.

Lotfi A. Zadeh. Fuzzy sets. Information and Control 8 (1965), 338--353.