

IMPROVEMENTS OF THE FUZZY CROSS-CLUSTERING ALGORITHM

HORIA F. POP

Abstract. The aim of this paper is to introduce a way to improve the behaviour of the Fuzzy Hierarchical Cross-Clustering Algorithm. The Simultaneous Fuzzy n -Means Algorithm produces fuzzy n -partitions of objects and characteristics, but it does not offer a modality to associate each fuzzy set of objects to a fuzzy set of characteristics. Based on our previous experience we are able to propose such an association. For a study of the behaviour of the Fuzzy Hierarchical Cross-Clustering Algorithm using both the original and the Improved Simultaneous Fuzzy n -Means Algorithm, please see [7].

1. Introduction

The fuzzy sets theory, developed by Zadeh [9] allows an object to belong to many clusters with different membership degrees. The membership degree of an object to a certain class is supposed to be gradual rather than abrupt (either 0 or 1), revealing a basis for considering uncertainty and imprecision.

There are two opposite approaches to hierarchical clustering, namely *agglomerative* and *divisive* procedures. An agglomerative hierarchical classification places each object in its own cluster and gradually merges the clusters into larger and larger clusters until all objects are in a single cluster. The divisive hierarchical clustering reverses the process by starting with all the objects in a single cluster and subdividing it into smaller ones until, finally, each object is in a cluster of its own. The number of clusters to be generated may be either specified in advance or optimized by the algorithm itself according to certain criteria.

In the present paper we will approach the fuzzy hierarchical cross-classification algorithm (see also [3, 8]). We will show a problem in the algorithm, and we will propose a solution for it. In order to demonstrate the quality of our improvement, we will do a comparative study based on the behaviour both of the original fuzzy

Received by the editors: January 5, 1998.

1991 *Mathematics Subject Classification.* 68H30, 68T10.

1991 *CR Categories and Descriptors.* I.5.1 [Pattern Recognition]: Models – Fuzzy set, I.5.3 [Pattern Recognition]: Clustering – Algorithms, Similarity measures.

hierarchical algorithm and of our improved version of it. For other interesting results in fuzzy clustering see [5, 6, 4].

2. Hierarchical cross-classification

In certain situations the number of characteristics is very large. The design of a hierarchical classifier may be simplified if at every node is used only a small subset of the characteristics, enough for the classification decision at that node. So, at every step of the hierarchical classification process we determine a fuzzy partition of a certain class and the relevant characteristics for each of the subclasses obtained. Thus, it is necessary to classify both the objects and the characteristics. This classification procedure will be called

cross-classification (or simultaneous classification).

In what follows we will recall [3] a method which allows us to obtain an objects hierarchy and a characteristics hierarchy, so that the two hierarchical classifications should correspond to each other. This method is

iterative and the classification is done alternatively on the data set and on the characteristics set, until we will obtain two "compatible" classifications.

2.1. Fuzzy substructure of a fuzzy set. In this section we will recall the so-called Generalized Fuzzy n -Means Algorithm [2, 3]. This algorithm is a generalization of the well-known Fuzzy n -Means Algorithm [1].

Let us consider a set of objects $X = \{x^1, \dots, x^p\} \subset \mathbb{R}^s$ and let C be a fuzzy set on X . We search for the cluster substructure of the fuzzy set C .

Let us suppose that the fuzzy partition corresponding to this substructure is $\{A_1, \dots, A_n\}$. We admit that each class A_i may be represented by a prototype L^i from the representation space, \mathbb{R}^s . If L^i is from X we may suppose that L^i has the greatest membership degree to A_i , that is:

$$(1) \quad A_i(L^i) = \max_{x \in X} A_i(x)$$

Let us denote by d a distance in the space \mathbb{R}^s . For example, we may consider the distance induced by the norm of the space.

The dissimilarity $D_i(x^j, L^i)$ between a point x^j and the prototype L^i is defined as the square distance to the class A_i and is interpreted as a measure of the inadequacy of the representation of the point x^j by the prototype L^i .

If L^i is not a point from the data set X , then we have from [2] that:

$$(2) \quad D_i(x^j, L^i) = (A_i(x^j))^2 d^2(x^j, L^i).$$

The inadequacy between the fuzzy partition P and its representation, $L = \{L^1, \dots, L^n\}$ is given by the following function:

$$(3) \quad J(P, L) = \sum_{i=1}^n \sum_{j=1}^p (A_i(x^j))^2 d^2(x^j, L^i)$$

$J(P, L)$ may also be interpreted as the representation error of P by L .

Let us observe that J is a criteria function of the type of square errors sum. The classification problem becomes the determination of the fuzzy partition P and its representation L for which the inadequacy $J(P, L)$ is minimal. Let us note that, intuitively, to minimize J means to give small membership degrees to A_i for those points in X for which the dissimilarity to the prototype L^i is large, and vice-versa.

If we admit that d is a distance induced by the norm, we may write

$$(4) \quad J(P, L) = \sum_{i=1}^n \sum_{j=1}^p (A_i(x^j))^2 \|x^j - L^i\|^2$$

If the norm is induced by the inner product, we have

$$\|x^j - L^i\|^2 = (x^j - L^i)^T M (x^j - L^i),$$

where M is a symmetrical and positively defined matrix. The transposing operation was denoted by T .

The criteria function becomes:

$$(5) \quad J(P, L) = \sum_{i=1}^n \sum_{j=1}^p (A_i(x^j))^2 (x^j - L^i)^T M (x^j - L^i)$$

Because an algorithm to obtain an exact solution of the problem (5) is not known, we will use an approximate method in order to determine a local solution. The minimum problem will be solved using an iterative (relaxation) method, where J is successively minimized with respect to P and L .

Supposing that L is given, the minimum of the function $J(\cdot, L)$ is obtained [2] for:

$$(6) \quad A_i(x^j) = \frac{C(x^j)}{\sum_{k=1}^n \frac{d^2(x^j, L^k)}{d^2(x^j, L^i)}}$$

if for all i , $d(x^j, L^i) \neq 0$, and, respectively, if for a certain x^j there is an L^i such that $d(x^j, L^i) = 0$, the memberships of x^j verify the condition

$$A_i(x^j) = 0, \forall i \text{ such that } d(x^j, L^i) \neq 0.$$

For a given P , the minimum of the function $J(P, \cdot)$ is obtained for:

$$(7) \quad L^i = \frac{\sum_{j=1}^p (A_i(x^j))^2 x^j}{\sum_{j=1}^p (A_i(x^j))^2}, i = 1, 2, \dots, n.$$

We observe [2] that L^i is the weighting center of the class A_i .

The iterative procedure for obtaining the cluster substructure of the fuzzy class C is called Generalized Fuzzy n-Means (GFNM) [2]. Essentially, the GFNM algorithm works with Picard iterations using the relations (6) and (7). The iterative process begins with an arbitrary initialization of the partition P . The process ends when two successive partitions are close enough. To measure the distance between two partitions, we will associate to each partition P a matrix Q with the dimensions $n \times p$. Q is named the representation matrix of the fuzzy partition P and is defined as:

$$(8) \quad Q_{ij} = A_i(x^j), i = 1, 2, \dots, n; j = 1, 2, \dots, p.$$

Considering that Q_1 and Q_2 are the representation matrices of the partitions P_1 and P_2 , we may define

$$(9) \quad d(P_1, P_2) = \|Q_1 - Q_2\|.$$

where $\|Q\| = \max_{i,j} |A_i(x^j)|$.

The process ends at the r -th iteration if

$$(10) \quad d(P_r, P_{r+1}) < \epsilon$$

where ϵ is an admissible error (usually, 10^{-5}).

For $C = X$ this procedure is the well known algorithm Fuzzy n-Means (FNM) [1].

2.2. Simultaneous Fuzzy n -Means Algorithm. Let $X = \{x^1, \dots, x^p\} \subset \mathbb{R}^d$ be the set of objects to be classified. A characteristic may be specified by its values corresponded to the p objects. So, we may say that $Y = \{y^1, \dots, y^d\} \subset \mathbb{R}^p$ is the set of characteristics. y_j^k is the value of the characteristic k with respect to the object j , so we may write $y_j^k = x_k^j$.

Let P be a fuzzy partition of the fuzzy set C of objects and Q be a fuzzy partition of the fuzzy set D of characteristics. The problem of the cross-classification (or simultaneous classification) is to determine the pair (P, Q) which optimizes a certain criterion function. By starting with an initial partition P^0 of C and an initial partition Q^0 of D , we will obtain a new partition P^1 . The pair

IMPROVEMENTS OF THE FUZZY CROSS-CLUSTERING ALGORITHM

(P^1, Q^0) allows us to determine a new partition Q^1 of the characteristics. The algorithm consists in producing a sequence (P^k, Q^k) of pairs of partitions, starting from the initial pair (P^0, Q^0) , in the following steps:

- (i): $(P^k, Q^k) \rightarrow (P^{k+1}, Q^k)$;
- (ii): $(P^{k+1}, Q^k) \rightarrow (P^{k+1}, Q^{k+1})$.

The rationale of the hierarchical cross-classification method [3] essentially supposes the splitting the sets X and Y in two subclasses. The obtained classes are alternatively divided in two subclasses, and so on. The two hierarchies may be represented by the same tree, having in each node a pair (C, D) , where C is a fuzzy set of objects and D is a fuzzy set of characteristics.

As a first step we propose ourselves to simultaneously determine the fuzzy partitions (as a particular case, the binary fuzzy partitions) of the classes C and D , so that the two partitions should be highly correlated.

With the Generalized Fuzzy n -Means algorithm, we will determine a fuzzy partition $P = \{A_1, \dots, A_n\}$ of the class C , using the original characteristics.

In order to classify the characteristics, we will compute their values for the classes $A_i, i = 1, \dots, n$. The value \bar{y}_i^k of the characteristic k with respect to the class A_i is defined as:

$$(11) \quad \bar{y}_i^k = \sum_{j=1}^p A_i(x^j) x_k^j, i = 1, \dots, n; k = 1, \dots, d.$$

Thus, from the original d p -dimensional characteristics we computed d new n -dimensional characteristics which are conditioned by the classes $A_i, i = 1, \dots, n$. We may admit that these new characteristics do not describe objects, but they characterize the classes A_i .

Let us consider now the set $\bar{Y} = \{\bar{y}^1, \dots, \bar{y}^d\}$ of the modified characteristics. We define the fuzzy set \bar{D} on \bar{Y} , given by

$$\bar{D}(\bar{y}^k) = D(y^k), k = 1, \dots, d.$$

The way the set \bar{Y} has been obtained lets us to conclude that if we will obtain an optimal partition of the fuzzy set D , this partition will be highly correlated to the optimal partition of the class C .

With the Generalized Fuzzy n -Means algorithm we will determine a fuzzy partition $Q = \{B_1, \dots, B_n\}$ of the class D , by using the characteristics given by the relation (11).

We may now characterize the objects in X with respect to the classes of properties $B_i, i = 1, \dots, n$. The value \bar{x}_i^j of the object j with respect to the class B_i is defined as:

$$(12) \quad \bar{x}_i^j = \sum_{k=1}^d B_i(\bar{y}^k) x_k^j, i = 1, \dots, n; k = 1, \dots, d.$$

Thus, from the original p d -dimensional objects we have computed p new n -dimensional objects, which correspond to the classes of characteristics $B_i, i = 1, \dots, n$.

Let us now consider the set $\bar{X} = \{\bar{x}^1, \dots, \bar{x}^p\}$ of the modified objects. We define the fuzzy set \bar{C} on \bar{X} , given by

$$\bar{C}(\bar{x}^j) = C(x^j), j = 1, \dots, p.$$

With the Generalized Fuzzy n -Means algorithm we will determine a fuzzy partition $P' = \{A'_1, \dots, A'_n\}$, of the class C by using the objects given by the relation (12).

The process continues until two successive partitions of objects (or of characteristics) are closed enough to each other. Thus, we have obtained **The Simultaneous Fuzzy n -Means Algorithm** (see [3]).

S1: Set $l = 0$. With the Generalized Fuzzy n -Means Algorithm we determine a fuzzy n -partition $P^{(l)}$ of the class C by using the initial objects.

S2: With the Generalized Fuzzy n -Means Algorithm we determine a fuzzy n -partition $Q^{(l)}$ of the class D by using the characteristics defined in (11).

S3: With the Generalized Fuzzy n -Means Algorithm we determine a fuzzy n -partition $P^{(l+1)}$ of the class C by using the objects defined in (12).

S4: If the partitions $P^{(l)}$ and $P^{(l+1)}$ are close enough, that is if

$$\|P^{(l+1)} - P^{(l)}\| < \epsilon,$$

where ϵ is a preset value, then **Stop**, else increase l by 1 and go to **S2**.

2.3. Associative Simultaneous Fuzzy n -Means (ASF) Algorithm. Let us remark that the Simultaneous Fuzzy n -Means Algorithm makes no explicit association of a fuzzy set A_i on X with a fuzzy set B_j on Y , i.e. what is the fuzzy set B_j that best describes the essential characteristics corresponding to the fuzzy set A_i . It only supposes that A_i is to be associated with B_i , and this is not always true, and thus it may sometimes produce wrong results. In this section we will correct this problem by proposing a better way of associating the fuzzy sets of objects and characteristics.

Let us have the set $X = \{x^1, \dots, x^n\} \subset \mathbb{R}^d$ of classification objects, and let Y be the correspondent set of

characteristics. Let us also consider the fuzzy sets C on X and D on Y .

Let us suppose that $P = \{A_1, \dots, A_n\}$ is the fuzzy n -partition of X and $Q = \{B_1, \dots, B_n\}$ is the fuzzy n -partition of Y produced by the Simultaneous Fuzzy n -Means Algorithm.

IMPROVEMENTS OF THE FUZZY CROSS-CLUSTERING ALGORITHM

Let us denote by S_n the set of all permutations on $\{1, \dots, n\}$. We wish to build that permutation $\sigma \in S_n$ which best associates the fuzzy set A_i with the fuzzy set $B_{\sigma(i)}$, for every $i = 1, \dots, n$.

Our aim is to build some function $J : S_n \rightarrow \mathbf{R}$ so that the optimal permutation σ is that which maximizes this function.

Let us consider the matrix $Z \in \mathbf{R}^{n,n}$ given by

$$(13) \quad z_{kl} = \sum_{j=1}^p \sum_{i=1}^d A_k(x^j) B_l(y^i) x_j^i, k, l = 1, \dots, n.$$

Let us remark the similarity between the way we compute the matrix Z in (13) and the way we computed the new objects and characteristics in relations (11) and (12).

The experience enables us to consider the function J as given by

$$(14) \quad J(\sigma) = \prod_{i=1}^n z_{i, \sigma(i)}.$$

Thus, supposing that the permutation σ maximizes the function J defined above, we will be able to associate the fuzzy set A_i with the fuzzy set $B_{\sigma(i)}$, $i = 1, \dots, n$. As we will see in the comparative study below, this association is more natural than the association of A_i with B_i , $i = 1, \dots, n$.

Based on these considerations we are able to introduce the following algorithm, that we have called **The Associative Simultaneous Fuzzy n -Means Algorithm (ASF)**.

- S1:** Set $l = 0$. With the Generalized Fuzzy n -Means Algorithm we determine a fuzzy n -partition $P^{(l)}$ of the class C by using the initial objects.
- S2:** With the Generalized Fuzzy n -Means Algorithm we determine a fuzzy n -partition $Q^{(l)}$ of the class D by using the characteristics defined in (11).
- S3:** With the Generalized Fuzzy n -Means Algorithm we determine a fuzzy n -partition $P^{(l+1)}$ of the class C by using the objects defined in (12).
- S4:** If the partitions $P^{(l)}$ and $P^{(l+1)}$ are close enough, that is if

$$\|P^{(l+1)} - P^{(l)}\| < \epsilon,$$

where ϵ is a preset value, then go to **S5**, else increase l by 1 and go to **S2**.

- S5:** Compute the permutation σ that maximizes the function J given in relation (14).

- S6:** Relabel the fuzzy sets B_i , so that $B_{\sigma(i)}$ becomes B_i , $i = 1, \dots, n$.

Let us remark now that, after the steps **S5** and **S6**, we are able to associate the fuzzy set A_i with the fuzzy set B_i , $i = 1, \dots, n$.

The single problem that needs to be detailed is the computation implied at the step **S5**. As we will see in the following section, this algorithm is intended to be used in a hierarchic procedure for obtaining a binary classification tree. Thus, the ASF algorithm will only be used for $n = 2$, and the problem of the computation at the step **S5** is trivial. Of course, for the general case where $n > 2$, the problem of this computation needs to be addressed separately, but, as we said, we are interested here in the hierarchic algorithm.

2.4. Fuzzy Hierarchical Cross-Classification Algorithm. In this section we will present three variants of the procedure of hierarchical cross-classification. As we will see, these variants are useful when approaching different kinds of problems related to objects-characteristics relationships.

Variant A. The method described below is the straightforward way of developing a hierarchical algorithm that should use at each node of the classification tree our ASF algorithm for $n = 2$.

We will first show the way of building the classification binary tree.

The nodes of the tree are labeled with a pair (C, D) , where C is a fuzzy set from a fuzzy partition of objects and D is a fuzzy set from a fuzzy partition of characteristics. The root node corresponds to the pair (X, Y) . In the first step the two sub-nodes (A_1, B_1) and respectively (A_2, B_2) will be computed by using the ASF algorithm. Of course, these two nodes will be effectively built only if the fuzzy partitions $\{A_1, A_2\}$ and $\{B_1, B_2\}$ describe real clusters.

For each of the terminal nodes of the tree we try to determine partitions having the form $\{A_1, A_2\}$ and $\{B_1, B_2\}$, by using the ASF algorithm, modified as we have mentioned before. In this way the binary classification tree is extended with two new nodes, (A_1, B_1) and (A_2, B_2) . The process continues until for any terminal node we are not able to determine a structure of real clusters, either for the set of objects, or for the set of characteristics. The final fuzzy partitions will contain the fuzzy sets corresponding to the terminal nodes of the binary classification tree.

This variant of the algorithm, called **FHCCA algorithm**, seems to be suitable for application where the important idea is to get most of the relationships between different classes of objects and different classes of characteristics.

Variant B. The classification binary tree is built in the same way as for the variant A of the algorithm.

However, we slightly change the way the ASF algorithm is applied.

Let us denote the final partitions obtained at a certain node of the classification tree by $P = \{A_1, \dots, A_n\}$ and $Q = \{B_1, \dots, B_n\}$.

When trying to build the fuzzy partitions corresponding to the fuzzy sets A_i and B_i , $i = 1, \dots, n$, with the help of the ASF algorithm, at the step S1 of the algorithm, instead of using the original objects we may use the objects defined in

the relation (12) and determined at the step S3 before the end of the algorithm that produced the fuzzy partitions P and Q .

This variant of the algorithm, called **FHCCB algorithm** is quite similar to the FHCCA algorithm. However, let us remark the way the information obtained at a certain node of the classification tree is used when working on the subsequent nodes. Thus, suitable for application where the important idea is to get most of the relationships between different classes of objects and different classes of characteristics.

Variant C. The classification binary tree is built in the same way as for the variants A and B of the algorithm.

This time, the change in how we apply the ASF algorithm is more substantial.

When trying to build the fuzzy partition at a certain node of the classification tree, at the step S3 of the ASF algorithm, instead of using the objects defined in the relation (12), we will use the original objects, as in step S1.

This variant of the algorithm, called **FHCCC algorithm** is very different from the FHCCA and FHCCB algorithms. Due to the way it works, it seems to be suitable for applications where we are interested not only in a fuzzy partition corresponding to the cluster substructure of the data set, but also in knowing which fuzzy class of characteristics is responsible for the separation of each fuzzy class of objects. Let us remark the way this behaviour is different from the behaviours of the FHCCA and FHCCB algorithms.

3. Concluding remarks

The output of a fuzzy algorithm includes not only a partition but also additional information in the form of membership values.

Moreover, the fuzzy cross-classification algorithm, with the three variants presented here, produces both a fuzzy objects partition and a fuzzy characteristics partition, "compatible" with the former. We introduced a way to associate each fuzzy set of objects to a fuzzy set of characteristics. We have comparatively analyzed the behaviour of the three variants of the hierarchical cross-classification algorithm: FHCCA, FHCCB and FHCCC algorithms. As we have seen in the sections 3.1 and 3.2, the ASF algorithm produced in this way works better than the original Simultaneous Fuzzy n -Means introduced in [3] and presented here.

Thus, the advantages of this algorithm (with its three variants) include the ability to observe on one hand, the fuzzy classes obtained and the relations between them, and on the other hand the characteristics corresponding to each final class of objects (and which have contributed to the separation of the respective class).

The analyzed examples [7] confirm that the three versions of the Fuzzy Cross Clustering algorithm introduced in this paper, namely FHCCA, FHCCB

and FHCCC approach different aspects in the data, and for this reason all three of them should be used as necessary.

References

- [1] BEZDEK, J. C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
- [2] DUMITRESCU, D. Hierarchical pattern classification. *Fuzzy Sets and Systems* 28 (1988), 145-162.
- [3] DUMITRESCU, D., POP, H. F., AND SÂRBU, C. Fuzzy hierarchical cross-classification of Greek muds. *J. Chem. Inf. Comp. Sci.* 35 (1995), 851-857.
- [4] POP, H. F. *Intelligent Systems in Classification Problems*. PhD thesis, "Babes-Bolyai" University, Faculty of Mathematics and Computer Science, Cluj-Napoca, 1995.
- [5] POP, H. F., DUMITRESCU, D., AND SÂRBU, C. A study of Roman pottery (terra sigillata) using hierarchical fuzzy clustering. *Anal. Chim. Acta* 310 (1995), 269-279.
- [6] POP, H. F., AND SÂRBU, C. A new fuzzy regression algorithm. *J. Anal. Chem.* 68 (1996), 771-778.
- [7] POP, H. F., AND SÂRBU, C. The fuzzy hierarchical cross-clustering algorithm. Improvements and comparative study. *J. Chem. Inf. Comp. Sci.* 37 (1997), 510-516.
- [8] SÂRBU, C., HOROWITZ, O., AND POP, H. F. A fuzzy cross-classification of the chemical elements, based both on their physical, chemical and structural features. *J. Chem. Inf. Comp. Sci.* 36 (1996), 1098-1108.
- [9] ZADEH, L. A. Fuzzy sets. *Inf. Control* 8 (1965), 338-353.

"BABEȘ-BOLYAI" UNIVERSITY, FACULTY OF MATHEMATICS AND INFORMATICS, RO-3400
 CLUJ-NAPOCA, ROMANIA

E-mail address: hfpop@cs.ubbcluj.ro