UDC 004.021
Original Scientific Paper

An Algorithm for Detecting the Principal Allotment among Fuzzy Clusters and Its Application as a Technique of Reduction of Analyzed Features Space Dimensionality

Dmitri A. Viattchenin *viattchenin@mail.ru*
National Academy of Sciences of Belarus
United Institute of Informatics Problems

Abstract

This paper describes a modification of a possibilistic clustering method based on the concept of allotment among fuzzy clusters. Basic ideas of the method are considered and the concept of a principal allotment among fuzzy clusters is introduced. The paper provides the description of the plan of the algorithm for detection principal allotment. An analysis of experimental results of the proposed algorithm's application to the Tamura's portrait data in comparison with the basic version of the algorithm and with the NERFCM-algorithm is carried out. A methodology of the algorithm's application to the dimensionality reduction problem is outlined and the application of the methodology is illustrated on the example of Anderson's Iris data in comparison with the result of principal component analysis. Preliminary conclusions are formulated also.
Keywords: possibilistic clustering, fuzzy tolerance, allotment among fuzzy clusters, typical point, membership degree, dimensionality reduction

1. Preliminaries

In general, cluster analysis refers to a spectrum of methods, which try to divide a set of objects $X = \{x_1, ..., x_n\}$ into subsets, called clusters, which are pairwise disjoint, all non empty and reproduce X via union. Heuristic methods, hierarchical methods, optimization methods and approximation methods are main approaches to the cluster analysis problem solving.

Clustering algorithms in general can also be divided into two types: hard versus fuzzy. Hard clustering assigns each object to exactly one cluster. In fuzzy clustering, a given pattern does not necessarily belong to only one cluster, but can have varying degrees of memberships to several clusters. Heuristic methods of fuzzy clustering, hierarchical methods of fuzzy clustering and optimization methods of fuzzy clustering were proposed by different researchers, but the most widely used is the FCM-algorithm. Conceived by Dunn [5] and generalized by Bezdek [2], this family of algorithms is based on iterative optimization of a fuzzy objective function. Many fuzzy clustering algorithms are proposed by other researchers. Moreover, a possibilistic approach to clustering was proposed by Krishnapuram and Keller [8]. A concept of possibilistic partition is a basis of possibilistic clustering methods and the possibilistic membership values μ_{li}, $1 \le l \le c$, $1 \le i \le n$ can be interpreted as the values of typicality degree. The possibilistic approach to clustering can be considered as a way in the optimization approach in fuzzy clustering because all methods of possibilistic clustering are objective function-based methods. Different fuzzy clustering algorithms are described by Sato, Sato and Jain [9] in detail.

The most widespread approach in fuzzy clustering is the optimization approach. However, heuristic algorithms of fuzzy clustering display high level of essential clarity and low level of a complexity. Some heuristic clustering algorithms are based on a definition of a cluster concept and the aim of these algorithms is cluster detection in conform to a given definition. These algorithms are called direct clustering algorithms [13]. Direct heuristic algorithms of fuzzy clustering are simple and very effective in many cases. The algorithm of Couturier and Fioleau [4] is a very good illustration for these characterizations.

An outline for a new heuristic method of fuzzy clustering was presented by Viattchenin [12], where concepts of fuzzy α-cluster and allotment among fuzzy α-clusters were introduced and a basic version of direct fuzzy clustering algorithm was described and the version was called the AFC-algorithm. The basic version of direct fuzzy clustering algorithm requires that the number c of fuzzy α-clusters be fixed. The allotment of elements of the set of classified objects among fuzzy clusters can be considered as a special case of possibilistic partition. These facts were demonstrated in [14], [15]. That is why the basic version of the algorithm, which is described in [12], can be considered as a direct algorithm of possibilistic clustering and the algorithm can be called the $D\!-\!AFC(c)$-algorithm. The $D\!-\!AFC(c)$-algorithm is a basis for other clustering algorithms [13], [14].

The main goal of this paper is a detailed consideration of a modification of the $D\!-\!AFC(c)$-algorithm, which is oriented at detection of the unknown least number of compact and well-separated fuzzy clusters. For this purpose, the concept of the principal allotment among fuzzy α-clusters is introduced in the paper. So, the modification can be called as the $D\!-\!PAFC$-algorithm. The contents of this paper are following: in the second section basic concepts of the method considered, the concept of a principal allotment is introduced and the general plan of the $D\!-\!PAFC$-algorithm is proposed, in the third section a numerical example of application of the $D\!-\!PAFC$-algorithm to the Tamura's portrait data in comparison with the $D\!-\!AFC(c)$-algorithm and with the well-known relational fuzzy clustering $NERFCM$-algorithm of Hathaway and Bezdek [7] is given, in the fourth section a methodology of the $D\!-\!PAFC$-algorithm's application to the solving of the problem of reduction of analyzed features space dimensionality is described and illustrated on the Anderson's Iris data example in comparison with the result of conventional principal component analysis, in the fifth section some final remarks are stated.

2. General Plan of the D-PAFC-algorithm

Let us remind the basic concepts of the possibilistic clustering method based on the concept of allotment among fuzzy clusters, which was proposed in [12]. The structure of the set of objects can be described by some fuzzy tolerance, that is – a fuzzy binary symmetric reflexive intransitive relation. The concept of fuzzy tolerance is the basis for the concept of fuzzy α-cluster. That is why definition of fuzzy tolerance must be considered in the first place.

Let $X = \{x_1, ..., x_n\}$ be the initial set of elements and $T : X \times X \to [0,1]$ some binary fuzzy relation on $X = \{x_1, ..., x_n\}$ with $\mu_T(x_i, x_j) \in [0,1]$, $\forall x_i, x_j \in X$ being its membership function.

Definition 1. *Fuzzy tolerance is the fuzzy binary intransitive relation which possesses the symmetricity property*

$$\mu_T(x_i, x_j) = \mu_T(x_j, x_i), \forall x_i, x_j \in X, \quad (1)$$

and the reflexivity property

$$\mu_T(x_i, x_i) = 1, \forall x_i \in X. \quad (2)$$

Let us consider basic concepts of the method. Let $X = \{x_1, ..., x_n\}$ be the initial set of objects. Let T be a fuzzy tolerance on X and α be α-level value of T, $\alpha \in (0,1]$. Columns or lines of the fuzzy tolerance matrix T are fuzzy sets $\{A^1, ..., A^n\}$ on X. Let $\{A^1, ..., A^n\}$ be fuzzy sets on X, which are generated by a fuzzy tolerance T.

Definition 2. *The α-level fuzzy set $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha\}$, $x_i \in X$, $l \in \{1, \ldots, n\}$ is fuzzy α-cluster or, simply, fuzzy cluster. So, $A^l_{(\alpha)} \subseteq A^l$, $\alpha \in (0,1]$, $A^l \in \{A^1, \ldots, A^n\}$ and μ_{li}*

is the membership degree of the element $x_i \in X$ *for some fuzzy cluster* $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1,\dots,n\}$. *Value of* α *is the tolerance threshold of fuzzy clusters elements.*

The membership degree of the element $x_i \in X$ for some fuzzy cluster $A^l_{(\alpha)}, \alpha \in (0,1], l \in [1,n]$ can be defined as a

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), x_i \in A^l_\alpha \\ 0, \qquad otherwise \end{cases}, \tag{3}$$

where an α-level $A^l_\alpha = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$ of a fuzzy set A^l is the support of the fuzzy cluster $A^l_{(\alpha)}$. So, condition $A^l_\alpha = Supp(A^l_{(\alpha)})$ is met for each fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$.

In other words, if columns or lines of fuzzy tolerance T matrix are fuzzy sets $\{A^1,\dots,A^n\}$ on X then fuzzy clusters $\{A^1_{(\alpha)},\dots,A^n_{(\alpha)}\}$ are fuzzy subsets of fuzzy sets $\{A^1,\dots,A^n\}$ for some value $\alpha, \alpha \in (0,1]$. The value zero for a fuzzy set membership function is equivalent to non-belonging of an element to a fuzzy set. That is why values of tolerance threshold α are considered in the interval $(0,1]$. So, a fuzzy cluster can be understood as some fuzzy subset originated by fuzzy tolerance relation stipulating that the similarity degree of the fuzzy subset elements is not less than some threshold value. In other words, the value of a membership function of each element of the fuzzy cluster is the degree of similarity of the object to some typical object of fuzzy cluster. So, membership degree can be interpreted as a degree of typicality of an element to a fuzzy cluster. Moreover, membership degree defines a possibility distribution function for some fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$. The fact was demonstrated by Viattchenin [15] and the possibility distribution function is denoted by $\pi_l(x_i)$.

Definition 3. *If* T *is a fuzzy tolerance on* X, *where* X *is the set of elements, and* $\{A^1_{(\alpha)},\dots,A^n_{(\alpha)}\}$ *is the family of fuzzy clusters for some* α, *then the point* $\tau^l_e \in A^l_\alpha$, *for which*

$$\tau^l_e = \arg\max_{x_i} \mu_{li}, \forall x_i \in A^l_\alpha \tag{4}$$

is called a typical point of the fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$.

Obviously, a typical point of a fuzzy cluster does not depend on the value of tolerance threshold. Moreover, a fuzzy cluster can have several typical points. That is why symbol e is the index of the typical point.

Definition 4. *Let* $R^\alpha_z(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, 2 \leq c \leq n\}$ *be a family of fuzzy clusters for some value of tolerance threshold* $\alpha \in (0,1]$, *which are generated by some fuzzy tolerance* T *on the initial set of elements* $X = \{x_1,\dots,x_n\}$. *If condition*

$$\sum_{l=1}^{c} \mu_{li} > 0, \quad \forall x_i \in X \tag{5}$$

is met for all fuzzy clusters $A^l_{(\alpha)}, l = \overline{1,c}, c \leq n$, *then the family is the allotment of elements of the set* $X = \{x_1,\dots,x_n\}$ *among fuzzy clusters* $\{A^l_{(\alpha)} \mid l = \overline{1,c}, 2 \leq c \leq n\}$ *for some value of the tolerance threshold* α.

It should be noted that several allotments $R^\alpha_z(X)$ can exist for some tolerance threshold α. That is why symbol z is the index of an allotment. The condition (5) requires that every object $x_i, i = \overline{1,n}$ must be assigned to at least one fuzzy cluster $A^l_{(\alpha)}, l = \overline{1,c}, c \leq n$ with the membership degree higher than zero. The condition $2 \leq c \leq n$ requires that the number of fuzzy clusters in $R^\alpha_z(X)$ must be more than two. Otherwise, the unique fuzzy cluster will contain all objects, possibly with different positive membership degrees. The definition of the allotment among fuzzy clusters (5) is similar to the definition of the possibilistic partition [8]. That is why the allotment among fuzzy clusters can be considered as the possibilistic partition and fuzzy clusters in the sense of definition 2 are elements of the possibilistic partition. However, the concept of allotment will be used in further considerations.

Definition 5. *Allotment* $R_I^\alpha(X)=\{A_{(\alpha)}^l \mid l=\overline{1,n}\}$ *of the set of objects among* n *fuzzy clusters for some tolerance threshold* α *is the initial allotment of the set* $X=\{x_1,\ldots,x_n\}$.

In other words, if initial data are represented by a matrix of some fuzzy T then lines or columns of the matrix are fuzzy sets $A^l \subseteq X, l=\overline{1,n}$ and level fuzzy sets $A_{(\alpha)}^l, l=\overline{1,n}, \alpha\in(0,1]$ are fuzzy clusters. These fuzzy clusters constitute an initial allotment for some tolerance threshold α and they can be considered as clustering components.

Thus, the problem of cluster analysis can be defined in general as the problem of discovering the unique allotment $R^*(X)$, resulting from the classification process, which corresponds to either most natural allocation of objects among fuzzy clusters or to the researcher's opinion about classification. In the first case, the number of fuzzy clusters c is not fixed. In the second case, the researcher's opinion determines the kind of the allotment sought and the number of fuzzy clusters c can be fixed.

If some allotment $R_z^\alpha(X)=\{A_{(\alpha)}^l \mid l=\overline{1,c}, c\le n\}$ corresponds to the formulation of a concrete problem, then this allotment is an adequate allotment. In particular, if condition

$$\bigcup_{l=1}^{c} A_\alpha^l = X, \tag{6}$$

and condition

$$card(A_\alpha^l \cap A_\alpha^m)=0, \forall A_{(\alpha)}^l, A_{(\alpha)}^m, l\ne m, \alpha\in(0,1], \tag{7}$$

are met for all fuzzy clusters $A_{(\alpha)}^l, l=\overline{1,c}$ of some allotment $R_z^\alpha(X)=\{A_{(\alpha)}^l \mid l=\overline{1,c}, c\le n\}$ then the allotment is the allotment among fully separate fuzzy clusters.

However, fuzzy clusters in the sense of definition 2 can have an intersection area. If the intersection area of any pair of different fuzzy cluster is an empty set, then conditions (6) and (7) are met and fuzzy clusters are called fully separate fuzzy clusters. Otherwise, fuzzy clusters are called particularly separate fuzzy clusters and $w=\{0,\ldots,n\}$ is the maximum number of elements in the intersection area of different fuzzy clusters. Obviously, for $w=0$ fuzzy clusters are fully separate fuzzy clusters. So, the conditions (6) and (7) can be generalized for a case of particularly separate fuzzy clusters. Condition

$$\sum_{l=1}^{c} card(A_\alpha^l)\ge card(X), \forall A_{(\alpha)}^l \in R_z^\alpha(X), \alpha\in(0,1], card(R_z^\alpha(X))=c, \tag{8}$$

and condition

$$card(A_\alpha^l \cap A_\alpha^m)\le w, \forall A_{(\alpha)}^l, A_{(\alpha)}^m, l\ne m, \alpha\in(0,1], \tag{9}$$

are generalizations of conditions (6) and (7). Obviously, if $w=0$ in conditions (8) and (9) then conditions (6) and (7) are met.

The adequate allotment $R_z^\alpha(X)$ for some value of tolerance threshold α, $\alpha\in(0,1]$ is a family of fuzzy clusters which are elements of the initial allotment $R_I^\alpha(X)$ for the value of α and the family of fuzzy clusters should satisfy either the conditions (6) and (7) or the conditions (8) and (9). So, the construction of adequate allotments $R_z^\alpha(X)=\{A_{(\alpha)}^l \mid l=\overline{1,c}, c\le n, \alpha\in(0,1]\}$ for every α is a trivial problem of combinatorics. Several adequate allotments can exist. Thus, the problem consists in the selection of the unique adequate allotment $R^*(X)$ from the set B of adequate allotments, $B=\{R_z^\alpha(X)\}$, which is the class of possible solutions of the concrete classification problem and $B=\{R_z^\alpha(X)\}$ depends on the parameters the classification problem. The selection of the unique adequate allotment $R^*(X)$ from the set $B = \{R_z^\alpha(X)\}$ of adequate allotments must be made on the basis of evaluation of allotments. The criterion

$$F_1(R_z^\alpha(X),\alpha) = \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c\,, \tag{10}$$

where c is the number of fuzzy clusters in the allotment $R_z^\alpha(X)$ and $n_l = card(A_\alpha^l), A_{(\alpha)}^l \in R_z^\alpha(X)$ is the number of elements in the support of the fuzzy cluster $A_{(\alpha)}^l$, can be used for evaluation of allotments. Maximum of criterion (10) corresponds to the best allotment of objects among c fuzzy clusters. So, the classification problem can be characterized formally as determination of the solution $R^*(X)$ satisfying

$$R^*(X) = \arg \max_{R_z^\alpha(X) \in B(c)} F_1(R_z^\alpha(X),\alpha)\,, \tag{11}$$

where $B = \{R_z^\alpha(X)\}$ is the set of adequate allotments corresponding to the formulation of a concrete classification problem. The criterion (10) can be considered as the average total membership of objects in fuzzy clusters of the allotment $R_z^\alpha(X)$ minus $\alpha \cdot c$. The quantity $\alpha \cdot c$ regularizes with respect to the number of clusters c in the $R_z^\alpha(X)$.

Detection of an unknown minimal number of compact and well-separated fuzzy clusters can be considered as the aim of classification in some situations. So, the following concept was introduced by Viattchenin [13].

Definition 6. *Allotment* $R_P^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}, \alpha \in (0,1]\}$ *of the set of objects among the minimal number* c, $2 \le c \le n$ *of fully separate fuzzy clusters for some tolerance threshold* α, $\alpha \in (0,1]$ *is the principal allotment of the set* $X = \{x_1,...,x_n\}$.

So, principal allotment must satisfy conditions (6) and (7). There is a five-step procedure of classification:

1. Calculate α-level values of the fuzzy tolerance T and construct the sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \le 1$ of α-levels; let $\ell := 1$;
2. Construct the initial allotment $R_I^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,n}\}$, $\alpha = \alpha_\ell$ for the value α_ℓ from the sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \le 1$;
3. The following condition is checked:
   if for some fuzzy cluster $A_{(\alpha)}^l \in R_I^\alpha(X), \alpha = \alpha_\ell$ the condition $card(A_{(\alpha)}^l) = n$ is met
   then let $\ell := \ell + 1$ and go to step 2
   else construct the allotments, which satisfy conditions (6) and (7);
4. The following condition is checked:
   if for α_ℓ allotments $R_z^\alpha(X)$ satisfying conditions (6) and (7) are not constructed
   then let $\ell := \ell + 1$ and go to step 2
   else construct the class of possible solutions of the classification problem $B^{\alpha_\ell} = \{R_z^\alpha(X)\}, \alpha = \alpha_\ell$ for α_ℓ;
5. The following condition is checked:
   if condition $card(B^{\alpha_\ell}) > 1$ is met
   then calculate the value of the criterion $F(R_z^\alpha(X),\alpha)$ for every allotment $R_z^\alpha(X) \in B^{\alpha_\ell}$
   and the result of the classification $R^*(X)$ must be constructed as follows:

if for some unique allotment $R_z^\alpha(X) \in B^{\alpha_\ell}$ the condition (11) is met
then the allotment is a solution $R^*(X)$ of the classification problem;
else **if** condition $card(B^{\alpha_\ell}) = 1$ is met
then the unique allotment $R_z^\alpha(X) \in B^{\alpha_\ell}$ is a solution $R^*(X)$ of the classification problem.

The principal allotment $R_P^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}, \alpha \in (0,1]\}$ among the unknown least number of fully separate fuzzy clusters and the corresponding value of tolerance threshold α are results of classification.

3. A Numerical Example

Let us consider an application of the proposed $D-PAFC$ -algorithm to the classification problem for the following illustrative example. The problem of classification of family portraits coming from three families was considered by Tamura, Higuchi and Tanaka [11].

The number of portraits was equal to 16 and the real portrait assignment among three classes is presented in Figure 1.

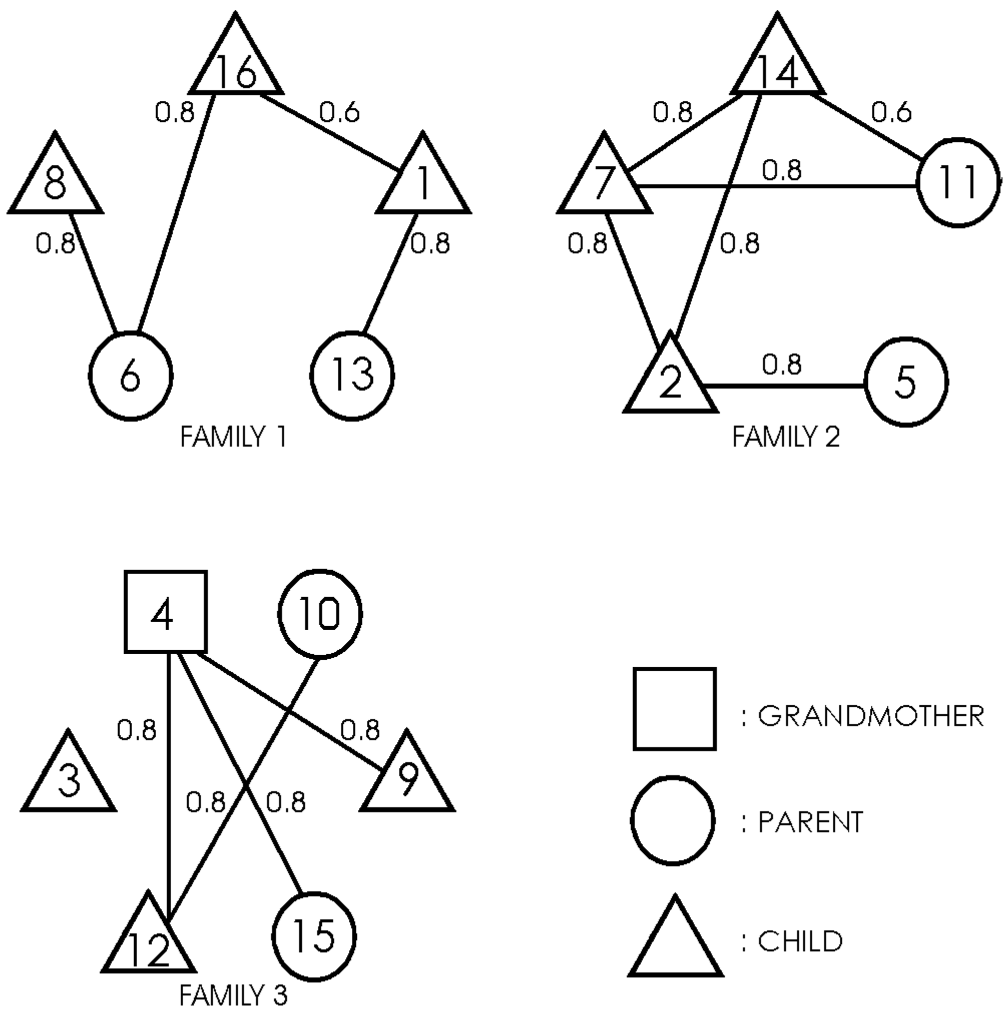

Figure 1: Real portrait classification.

The subjective similarities assigned to the individual pairs of portraits collected in the tabular format are presented in Table 1.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1.0															
2	0.0	1.0														
3	0.0	0.0	1.0													
4	0.0	0.0	0.4	1.0												
5	0.0	0.8	0.0	0.0	1.0											
6	0.5	0.0	0.2	0.2	0.0	1.0										
7	0.0	0.8	0.0	0.0	0.4	0.0	1.0									
8	0.4	0.2	0.2	0.5	0.0	0.8	0.0	1.0								
9	0.0	0.4	0.0	0.8	0.4	0.2	0.4	0.0	1.0							
10	0.0	0.0	0.2	0.2	0.0	0.0	0.2	0.0	0.2	1.0						
11	0.0	0.5	0.2	0.2	0.0	0.0	0.8	0.0	0.4	0.2	1.0					
12	0.0	0.0	0.2	0.8	0.0	0.0	0.0	0.0	0.4	0.8	0.0	1.0				
13	0.8	0.0	0.2	0.4	0.0	0.4	0.0	0.4	0.0	0.0	0.0	0.0	1.0			
14	0.0	0.8	0.0	0.2	0.4	0.0	0.8	0.0	0.2	0.2	0.6	0.0	0.0	1.0		
15	0.0	0.0	0.4	0.8	0.0	0.2	0.0	0.0	0.2	0.0	0.0	0.2	0.2	0.0	1.0	
16	0.6	0.0	0.0	0.2	0.2	0.8	0.0	0.4	0.0	0.0	0.0	0.0	0.4	0.2	0.0	1.0

Table 1. The matrix of subjective similarities.

In fact, the matrix of subjective similarities is the matrix of a fuzzy tolerance. That is why $D-PAFC$ -algorithm can be applied to the matrix directly. The application of the $D-PAFC$ -algorithm to the classification problem was made in comparison with the basic version of the algorithm and the $NERFCM$ -algorithm of fuzzy clustering [7] for the number of classes $c=3$. In order to compare the proposed algorithm with the well-known relational fuzzy clustering $NERFCM$ -algorithm, we transformed the initial matrix into a dissimilarity matrix by complementing the relationship degrees. The membership values originating from all experiments are presented in Table 2.

Numbers of objects, i	Membership values									
	From the $NERFCM$ -algorithm			From the $D-AFC(c)$ -algorithm			From the $D-PAFC$ -algorithm			
	u_{1i}	u_{2i}	u_{3i}	μ_{1i}	μ_{2i}	μ_{3i}	μ_{1i}	μ_{2i}	μ_{3i}	μ_{4i}
1	0.64	0.17	0.19	1.0	0.0	0.0	0.5	0.0	0.0	0.0
2	0.10	0.79	0.11	0.0	1.0	0.0	0.0	1.0	0.0	0.0
3	0.32	0.26	0.42	0.0	0.0	0.2	0.0	0.0	0.4	0.0
4	0.08	0.06	0.86	0.0	0.0	0.8	0.0	0.0	0.8	0.0
5	0.25	0.49	0.26	0.0	0.8	0.0	0.0	0.8	0.0	0.0
6	0.72	0.12	0.16	0.5	0.0	0.0	1.0	0.0	0.0	0.0
7	0.09	0.81	0.10	0.0	0.8	0.0	0.0	0.8	0.0	0.0
8	0.56	0.19	0.25	0.4	0.2	0.0	0.8	0.0	0.0	0.0
9	0.21	0.30	0.49	0.0	0.4	0.4	0.0	0.4	0.0	0.0
10	0.27	0.29	0.44	0.0	0.0	0.8	0.0	0.0	0.0	1.0
11	0.20	0.57	0.23	0.0	0.5	0.0	0.0	0.5	0.0	0.0
12	0.18	0.17	0.65	0.0	0.0	1.0	0.0	0.0	0.0	0.8
13	0.55	0.19	0.26	0.8	0.0	0.0	0.4	0.0	0.0	0.0
14	0.13	0.73	0.14	0.0	0.8	0.0	0.0	0.8	0.0	0.0
15	0.25	0.21	0.54	0.0	0.0	0.2	0.0	0.0	1.0	0.0
16	0.65	0.16	0.19	0.6	0.0	0.0	0.8	0.0	0.0	0.0

Table 2. The results of applications of clustering algorithms.

The fuzzy c-partition $P(X)$ is produced by the $NERFCM$-algorithm. So, condition $\sum_{l=1}^{c} u_{li} = 1$ is met for the membership values u_{li}, $0 \le u_{li} \le 1$, $1 \le l \le c$, $1 \le i \le n$ originating from the $NERFCM$-algorithm. If the maximum memberships rule is applied to the matrix $P_{3\times16} = [u_{li}]$, $l = 1,\ldots,c$, $i = 1,\ldots,n$ of the fuzzy c-partition $P(X)$, then the result of the $NERFCM$-algorithm application to the Tamura's data is similar to the real portraits classification.

By executing the $D\!-\!AFC(c)$-algorithm for three classes, we obtain the allotment $R^{*}(X)$ among particularly separate fuzzy clusters, which corresponds to the result, is received for the tolerance threshold $\alpha = 0.2$. The ninth element of the set of objects is belonging to the second class and to the third class and membership values are equal, $\mu_{29} = \mu_{39} = 0.4$. Obviously, the membership function obtained from the $D\!-\!AFC(c)$-algorithm is sharper than the membership function resulting from the $NERFCM$-algorithm.

After application of the $D\!-\!PAFC$-algorithm to the matrix of the initial data, the principal allotment $R_{P}^{0.4}(X)$ among four fuzzy clusters, which corresponds to the result, is received for the tolerance threshold $\alpha = 0.4$. By executing the $D\!-\!PAFC$-algorithm we obtain the following: the first class is composed of 5 elements, all belonging to Family 1; the second class is formed by 6 elements, where five elements correspond to Family 2 and one element corresponds to Family 3; the third class consists of 3 elements, all belonging to Family 3, and the fourth class contains 2 elements, all from Family 3. So, the union of the third and fourth classes is the class, which corresponds to Family 3 and there is one mistake of classification. The ninth element of the set of objects is the misclassified object.

The value of the membership function of the first fuzzy cluster is maximal for the sixth object. So, the sixth object is the typical point of the first fuzzy cluster. The membership value of the second object is equal one for the fuzzy cluster, which corresponds to the second class and the second object is the typical point of the second fuzzy cluster. The membership value of the fifteenth object is equal one for the third fuzzy cluster. That is why the fifteenth object is the typical point of the fuzzy cluster which corresponds to the third class. The value of the membership function of the fourth fuzzy cluster is maximal for the tenth object. So, the tenth object is the typical point of the fuzzy cluster which corresponds to the fourth class.

The matrix of the allotment can be illustrated by a diagram. Membership functions of four classes are presented in Figure 2.

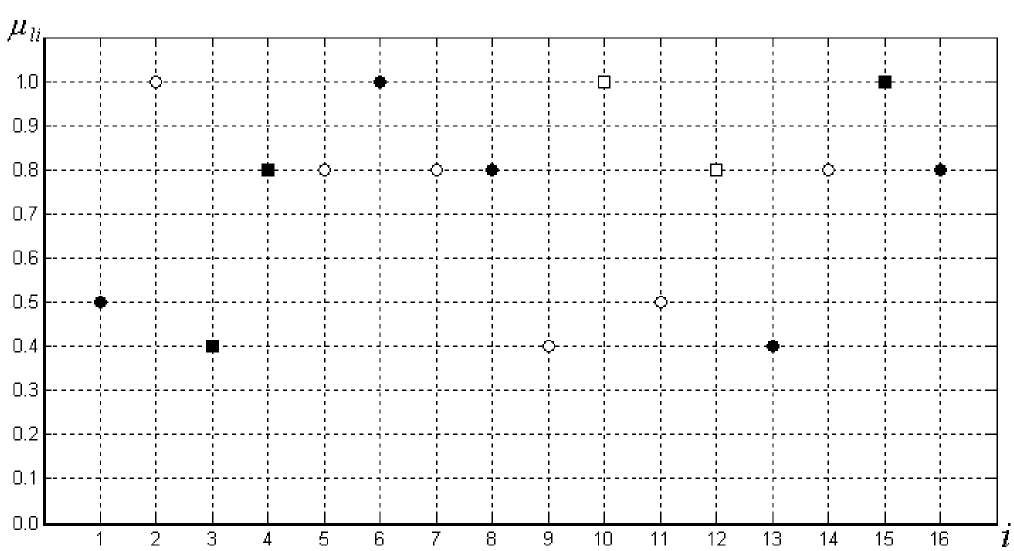

Figure 2. Membership functions of four classes obtained from the D-PAFC-algorithm.

Membership values of the first class are represented in Figure 2 by •, membership values of the second class are represented in Figure 2 by ○, membership values of the third class are represented in Figure 2 by ■, and membership values of the fourth class are represented in Figure 2 by □.

4. A Methodology of Dimensionality Reduction Based on the D-PAFC-algorithm

Problems of data visualization and reduction of the analyzed feature space dimensionality are very important in the process of data analysis. Feature selection is meant here to refer to the problem of dimensionality reduction of the data which initially contain a high number of features. The selection aims to choose the minimal number of the original features which still contain the information essential for discovering of substructure in the data, while reducing the computational complexity imposed by using many features. Different feature selection methods were briefly described by Ghazavi and Liao in [6]. An application of fuzzy clustering methods to the problem of feature selection was outlined in [3].

The $D\!-\!PAFC$ -algorithm can be applied to solve the problem of reduction of feature space. The basic idea of the approach is that features can be classified and a typical point of each fuzzy cluster can be considered as an informative feature. In other words, the approach can be considered as a version of the method of extremal grouping of features. So, a methodology of solving the problem of reduction of feature space dimensionality can be described as follows:

1. The initial data are contained in the matrix of attributes $X_{m\times n}=[\hat{x}_i^t]$, $i=1,\ldots,n$, $t=1,\ldots,m$. So, the matrix of correlation coefficients $r_{m\times m}=[r(\hat{x}^t,\hat{x}^k)]$, $t=1,\ldots,m$, $k=1,\ldots,m$ can be constructed as follows:
$$r(\hat{x}^t,\hat{x}^k)=\frac{c_{tk}}{s^t s^k}, \tag{12}$$
where $c_{tk}=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{x}_i^t-\bar{\hat{x}}^t\right)\left(\hat{x}_i^k-\bar{\hat{x}}^k\right)$, $(s^t)^2=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{x}_i^t-\bar{\hat{x}}^t\right)^2$, and $\bar{\hat{x}}^t=\frac{1}{n}\sum_{i=1}^{n}\hat{x}_i^t$.
2. The matrix of correlation coefficients for the data can be normalized as follows:
$$\tilde{r}(\hat{x}^t,\hat{x}^k)=\frac{\left(r(\hat{x}^t,\hat{x}^k)-\min_{t,k} r(\hat{x}^t,\hat{x}^k)\right)}{\max_{t,k} r(\hat{x}^t,\hat{x}^k)-\min_{t,k} r(\hat{x}^t,\hat{x}^k)}. \tag{13}$$
So, the matrix of correlation coefficients after normalization can be treated as a matrix of fuzzy tolerance relation. The $D\!-\!PAFC$ -algorithm can be applied directly to the matrix of normalized correlation coefficients.
3. Typical points of fuzzy clusters of the received principal allotment $R_P^\alpha(X)$ can be selected as most informative attributes.

An application of the methodology to data visualization can be illustrated on the Anderson’s Iris data example. The Anderson's Iris data set consists of the sepal length, sepal width, petal length and petal width for 150 irises [1]. The problem of classification is to classify the plants into three subspecies on the basis of this information. Let us consider the problem of most informative feature selection.

The Anderson’s Iris data forms the matrix of attributes $X_{4\times 150}=[\hat{x}_i^t]$, $i=1,\ldots,150$, $t=1,\ldots,4$, where the sepal length is denoted by $\hat{x}^1$, sepal width is denoted by $\hat{x}^2$, petal length is denoted by $\hat{x}^3$ and petal width is denoted by $\hat{x}^4$. A matrix of correlation coefficients

$r_{4\times 4} = [r(\hat{x}^t, \hat{x}^k)]$, $t = 1,\dots,4$, $k = 1,\dots,4$ can be constructed using formula (12). The matrix of correlation coefficients presented in Table 3.

$r(\hat{x}^t, \hat{x}^k)$	$\hat{x}^1$	$\hat{x}^2$	$\hat{x}^3$	$\hat{x}^4$
$\hat{x}^1$	1.0000	-0.1176	0.8718	0.8169
$\hat{x}^2$	-0.1176	1.0000	-0.4284	-0.3661
$\hat{x}^3$	0.8718	-0.4284	1.0000	0.9629
$\hat{x}^4$	0.8179	-0.3661	0.9629	1.0000

Table 3. The matrix of correlation coefficients for the Anderson's Iris data.

After application of the formula (13) to the matrix of correlation coefficients a matrix of fuzzy tolerance relation was constructed. The matrix is presented in Table 4.

$\widetilde{r}(\hat{x}^t, \hat{x}^k)$	$\hat{x}^1$	$\hat{x}^2$	$\hat{x}^3$	$\hat{x}^4$
$\hat{x}^1$	1.0000	0.2176	0.9102	0.8725
$\hat{x}^2$	0.2176	1.0000	0.0000	0.0436
$\hat{x}^3$	0.9102	0.0000	1.0000	0.9740
$\hat{x}^4$	0.8725	0.0436	0.9740	1.0000

Table 4. The matrix of correlation coefficients after normalization.

The $D-PAFC$-algorithm was applied directly to the matrix of normalized correlation coefficients. The principal allotment $R_P^{0.9102}(X)$ among two fuzzy clusters was obtained. Results of the application are presented in Table 5.

Class	Attributes			
	$\hat{x}^1$	$\hat{x}^2$	$\hat{x}^3$	$\hat{x}^4$
1	0.0000	1.0000	0.0000	0.0000
2	0.9102	0.0000	1.0000	0.9740

Table 5. The results of D-PAFC-algorithm application: the attribute assignment.

By executing the $D-PAFC$-algorithm we obtain two fuzzy clusters in the principal allotment $R_P^{0.9102}(X)$. The second feature $\hat{x}^2$ is the typical point of the first fuzzy cluster and the third feature $\hat{x}^3$ is the typical point of the second fuzzy cluster.

So, features $\hat{x}^2$ and $\hat{x}^3$ can be selected as most informative indexes and the two-dimensional projection of the Anderson's Iris data can be constructed. The projection is presented in Figure 3.

Two well-separated classes are visualized. The first class corresponds to Iris Setosa. The second class corresponds Iris Versicolor and Iris Virginica. Objects known to be Iris Setosa are represented by ■ in Figure 3, while those known to be Iris Versicolor are represented by ○ in Figure 3, and Iris Virginica are represented by $\times$. Obviously, the approach to the data visualization can be very useful in the exploratory data analysis.

Notable that the result is similar to the result, obtained from conventional principal component analysis [10]. An interpretation of the obtained principal components can be made on a basis of the factor loading. The factor loading is defined as a correlation coefficient between v-th principal component z_v and the t-th attribute $\hat{x}^t$, $t = 1,\dots,m$ as follows [10]:

$$f(z_v, \hat{x}^t) = \frac{\text{cov}\{z_v, \hat{x}^t\}}{\sqrt{V\{z_v\}V\{\hat{x}^t\}}}, \tag{14}$$

where $V\{z_v\}$ is variance of z_v, $V\{\hat{x}^t\}$ is variance of $\hat{x}^t$, and $\text{cov}\{z_v, \hat{x}^t\}$ is covariance between z_v and $\hat{x}^t$. $f(z_2, \hat{x}^t)$

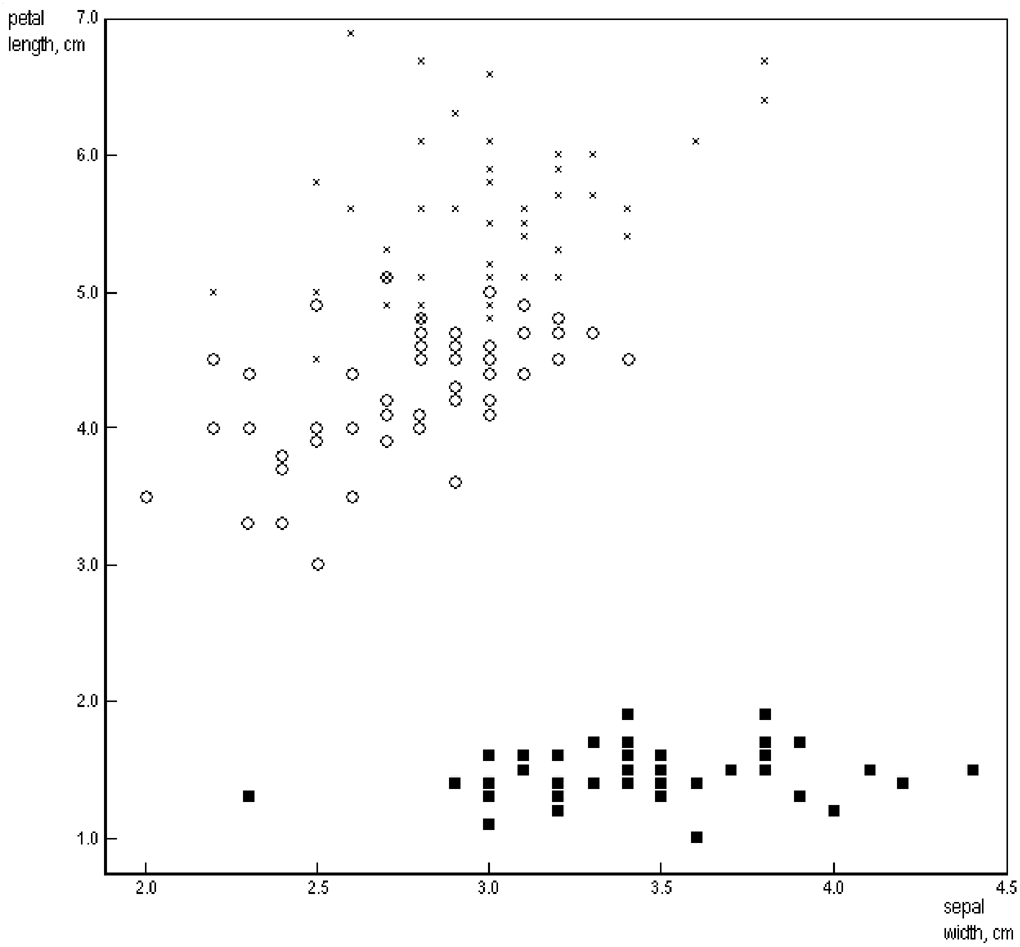

Figure 3. Two-dimensional projection of the Anderson's Iris data.

In Table 6, each value shows the value of the factor loadings (14) which can show the relationship between each principal component and each attribute.

Attributes	Principal components			
	z_1	z_2	z_3	z_4
$\hat{x}^1$	0.89	0.36	0.28	0.04
$\hat{x}^2$	-0.46	0.88	-0.09	-0.02
$\hat{x}^3$	0.99	0.02	-0.05	-0.12
$\hat{x}^4$	0.96	0.06	-0.24	0.08

Table 6. Factor loading in principal component analysis.

From the results, we can see how each component is explained by the attributes. This is related to the interpretation of each component.

In Table 6, the first principal component is mainly explained by the attributes, sepal length, petal length, and petal width. Moreover, we can see a high correlation between the second principal component and the attribute sepal width.

From the comparison between the results of Tables 5 and 6, we can see like results. In particular, values of the membership function of the first fuzzy cluster of the principal allotment $R_P^{0.9102}(X)$ can be interpreted as normalized values of the factor loadings $f(z_2, \hat{x}^t)$, $t = 1,\ldots,4$ in Table 6 and values of the membership function of the second fuzzy cluster of the principal allotment $R_P^{0.9102}(X)$ can be considered as normalized values of the factor loadings $f(z_1, \hat{x}^t)$, $t = 1,\ldots,4$ in Table 6.

5. Final Remarks

In conclusion it should be said that the concept of fuzzy cluster and allotment have an epistemological motivation. That is why the results of application of the possibilistic clustering method based on the allotment concept can be very well interpreted. Moreover, the possibilistic clustering method based on the allotment concept depends on the set of adequate allotments only. That is why the clustering results are stable.

The $D-PAFC$-algorithm of possibilistic clustering is proposed in the paper. The algorithm is based on the concept of the principal allotment among fuzzy clusters and an unknown minimal number of compact and well-separated fuzzy clusters is the result of classification. The result can be very useful in the exploratory data analysis. Moreover, the $D-PAFC$-algorithm does not depend on parameters and can be applied directly to the data given as the matrix of tolerance coefficients. This means that it can be used with the objects by attributes data, by choosing a suitable metric to measure similarity or it can be used in situations where objects by objects proximity data is available. The results of application of the $D-PAFC$-algorithm to the Tamura's portrait data show that the $D-PAFC$-algorithm is a precise and effective numerical procedure for solving classification problem.

A methodology of application of the $D-PAFC$-algorithm to the problem of reduction of feature space dimensionality and feature selection is proposed as well. The methodology can be considered as a version of the method of extremal grouping of features. The result of application of the methodology to the Anderson's Iris data shows that the methodology is a simple and effective tool for dimensionality reduction and can be applied for data visualization.

The $D-AFC(c)$-algorithm can be applied for a selection of subset of most appropriate weak fuzzy preference relations from the set of all weak fuzzy preference relations in group decision process [16]. However, the proposed $D-PAFC$-algorithm is seems as a more appropriate numerical procedure for solving the problem of discriminating fuzzy preference relations, because a problem of cluster validity can be solved immediately.

Acknowledgements

The extended version of the paper was presented as a lecture at the Department of Statistics of the Vienna University of Technology. I am grateful to Prof. Reinhard Viertl and his colleagues for their invitation and support. I would like to thank the anonymous referees and Prof. Neven Vrcek, Editor, for their constructive reviews and comments. I also thank Prof. Frank Klawonn for his interest in my investigations. I would like to thank Prof. Janusz Kacprzyk, Dr. Jan W. Owsinski, Prof. Alexander Tuzikov and Dr. Vassili Kovalev for fruitful discussions and their useful remarks during the manuscript preparation. I am grateful to Mr. Aliaksandr Damaratski for elaborating the experimental software.

References

[1] Anderson, E. The irises of the Gaspe Peninsula. *Bulletin of the American Iris Society*, 59:2-5, 1935.

[2] Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.

[3] Bezdek, J.C; Castelaz, P.F. Prototype classification and feature selection with fuzzy sets. *IEEE Transactions on Systems, Man, and Cybernetics,* 7(2):87-92, 1977.

[4] Couturier, A; Fioleau, B. Recognizing stable corporate groups: a fuzzy classification method. *Fuzzy Economic Review,* II(2): 35-45, 1997.

[5] Dunn, J.C. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics,* 3(3):32-57, 1974.

[6] Ghazavi, S.N; Liao, T.W. Medical data mining by fuzzy modeling with selected features. *Artificial Intelligence in Medicine,* 43(3):195-206, 2008.

[7] Hathaway, R.J; Bezdek, J.C. NERF C-means: non-Euclidean relational fuzzy clustering. *Pattern Recognition,* 27(3):429-437, 1994.

[8] Krishnapuram, R; Keller, J.M. A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98-110, 1993.

[9] Sato, M; Sato, Y; Jain, L.C. *Fuzzy Clustering Models and Applications*. Springer-Verlag, Heidelberg, 1997.

[10] Sato-Ilic, M; Jain, L.C. *Innovations in Fuzzy Clustering: Theory and Applications*. Springer-Verlag, Heidelberg, 2006.

[11] Tamura, S; Higuchi, S; Tanaka, K. Pattern classification based on fuzzy relations. *IEEE Transactions on Systems, Man, and Cybernetics,* 1(1):61-66, 1971.

[12] Viattchenin, D.A. A new heuristic algorithm of fuzzy clustering. *Control & Cybernetics*, 33(2):323-340, 2004.

[13] Viattchenin, D.A. Direct algorithms of fuzzy clustering based on the transitive closure operation and their application to outliers detection. *Artificial Intelligence,* 3:205-216, 2007. (in Russian)

[14] Viattchenin, D.A. An outline for a heuristic approach to possibilistic clustering. Working Paper WP-2-2007, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, 2007.

[15] Viattchenin, D.A. On possibilistic interpretation of membership values in fuzzy clustering method based on the allotment concept. *Proceedings of the Institute of Modern Knowledge*, 3:85-90, 2008. (in Russian)

[16] Viattchenin, D.A. Discriminating fuzzy preference relations based on heuristic possibilistic clustering. In Owsiński, J.W; Brüggemann, R, editors. *Multicriteria Ordering and Ranking: Partial Orders, Ambiguities and Applied Issues,* pages 197-213, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland, 2008.