

A NEW HEURISTIC POSSIBILISTIC CLUSTERING ALGORITHM FOR FEATURE SELECTION

Submitted: 14th September 2013; accepted 10th December 2013

Janusz Kacprzyk, Jan W. Owsinski, Dimitri A. Viattchenin

DOI 10.14313/JAMRIS_2-2014/18

Abstract: *The paper deals with the problem of selection of the most informative features. A new effective and efficient heuristic possibilistic clustering algorithm for feature selection is proposed. First, a brief description of basic concepts of the heuristic approach to possibilistic clustering is provided. A technique of initial data pre-processing is described and a fuzzy correlation measure is considered. The new algorithm is described and then illustrated on the well-known Iris data set benchmark and the results obtained are compared with those by using the conventional, well-known and widely employed method of principal component analysis (PCA). Conclusions and suggestions for future research are given.*

Keywords: *feature selection, fuzzy correlation measure, possibilistic clustering, heuristic possibilistic clustering, fuzzy cluster*

1. Introduction

1.1. Preliminary Remarks

The reduction of dimensionality of the feature space analyzed is very important a problem in data analysis. Feature selection is meant here as the dimensionality reduction of the feature space of data that has initially contained a high number of features. The purpose of the feature selection process is to choose a minimal number (subset) of the original set of features which still contain information that is essential for the discovering of substructure in the data while reducing the computational complexity implied by using a high number of features in the source problem formulation. Feature selection has been a fertile field of research, and has been development since the 1970s proving to be effective and efficient in removing irrelevant and/or redundant features, increasing the efficiency of learning, improving the learning performance characterized by, for instance, predictive accuracy, and enhancing the comprehensibility of results obtained. Many different feature selection methods have been proposed, cf. for example [1], [2], [3], and [4].

Fuzzy clustering methods can well be applied to solve the problem of feature selection. In particular, a combination of feature selection with feature weights and semi-supervised fuzzy clustering in machine learning is proposed by Kong and Wang [5]. On the other hand, a fuzzy feature selection method based on clustering was proposed by Chitsaz, Taheri,

and Katebi [6]. In the corresponding FACA-algorithm, each feature is assigned to different fuzzy clusters with different grades of membership. This comes from the basic underlying idea that each feature may belong not only to just one cluster, and it is much better to consider an association of each feature with other features in each cluster. Precise relations between features are therefore available during the selection of the most relevant features.

An extension of the FACA-algorithm is considered by Chitsaz, Taheri, Katebi and Jahromi [7] who have introduced four different techniques for implementing the stage of feature selection. For example, by applying the chi-square test, their approach considers the dependence of each feature on class labels in the process of feature selection.

1.2. A heuristic Approach to Possibilistic Clustering

The objective function based fuzzy clustering algorithms are the most widely employed methods in fuzzy clustering (cf. Bezdek, Keller, Krishnapuram and Pal [8]). Some heuristic clustering algorithms are based on the definition of the very concept of a cluster and the purpose of these algorithms is to find clusters according how they have been defined. Such algorithms are called direct classification (or clustering) algorithms (cf. Mandel [9]).

An outline of a new heuristic method of fuzzy clustering is presented by Viattchenin [10] who has considered a basic version of a direct clustering algorithm, while a version of such an algorithm, called the D-AFC(c)-algorithm, is presented in Viattchenin [11]. The D-AFC(c)-algorithm can be considered as a direct possibilistic clustering algorithm, as in [11]. The D-AFC(c)-algorithm has been shown there to be a basis for the family of other heuristic possibilistic clustering algorithms.

The direct heuristic possibilistic clustering algorithms can be divided into two types: relational and prototype-based. In particular, the family of direct relational heuristic possibilistic clustering algorithms includes:

- The D-AFC(c)-algorithm which is based on the construction of an allotment (to be defined later on in this paper) among an a priori given number c of partially separate fuzzy clusters [10];
- The D-AFC-PS(c)-algorithm which is based on the construction of an allotment among an a priori given number c of partially separate fuzzy clusters in the presence of labeled objects [11];

- The D-PAFC-algorithm which is based on the construction of an allotment among an unknown number of at least c fully separate fuzzy clusters [12].

It should be noted that the D-PAFC-algorithm can be applied to solve the problem of informative feature selection. The corresponding method was also proposed by Viattchenin [12].

On the other hand, the family of direct prototype-based heuristic possibilistic clustering algorithms, proposed by Viattchenin [13], includes:

- The D-AFC-TC-algorithm which is based on the construction of an allotment among an unknown number c of fully separate fuzzy clusters;
- The D-PAFC-TC-algorithm which is based on the construction of a principal allotment among an unknown minimal number of at least c fully separate fuzzy clusters;
- The D-AFC-TC(α)-algorithm which is based on the construction of an allotment among an unknown number c of fully separate fuzzy clusters with respect to a minimal value α of a tolerance threshold.

The unique allotment among an unknown number \tilde{n} of fuzzy clusters can be selected from the set of allotments depending on a tolerance threshold.

The main goal of this paper is to propose a new effective and efficient heuristic possibilistic clustering algorithm for solving the feature selection problem. The contents of this paper is as follows: in Section 2 some basic concepts of the heuristic approach to possibilistic clustering are briefly presented. In Section 3 a fuzzy correlation measure is proposed and a suitable method of data preprocessing is shown. In Section 4 the new heuristic possibilistic clustering algorithm is described. In Section 5 the new method is illustrated on the well-known benchmark example of the Iris data set, and a comparison with the results obtained by using the well-known and widely employed method of the principal components analysis (PCA) is presented. Conclusions are given in Section 6.

2. Outline of the Heuristic Possibilistic Clustering Algorithm

Let us remind the basic idea of, and the concepts related to the heuristic approach to possibilistic clustering. The concept of a fuzzy tolerance relation is the basis for the concept of a fuzzy α -cluster and that is why the definition of a fuzzy tolerance relation must be considered in the first place.

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

A *fuzzy tolerance relation* is a fuzzy binary relation which is symmetric, i.e.

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

and reflexive, i.e.

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

A *fuzzy similarity relation* S is a fuzzy binary relation which is symmetric (1), reflexive (2), and (max-min)-transitive, i.e.:

$$\begin{aligned} \mu_S(x_i, x_k) &\geq \bigvee_{x_j \in X} (\mu_S(x_i, x_j) \wedge \mu_S(x_j, x_k)), \\ \forall x_i, x_j, x_k &\in X. \end{aligned} \quad (3)$$

Let some fuzzy binary relation be represented by a matrix R of size $n \times n$, and let us define

$$R^1 = R, \quad R^n = R^{n-1} \circ R, \quad n = 2, 3, \dots \quad (4)$$

Now, the *transitive closure* of a fuzzy relation R is the fuzzy binary relation \bar{R} defined by

$$\bar{R} = R^1 \cup R^2 \cup \dots \cup R^n, \quad (5)$$

where the operation \cup for two fuzzy relations R^d and R^g is defined as

$$\begin{aligned} \mu_{R^d \cup R^g}(x_i, x_j) &= \mu_{R^d}(x_i, x_j) \vee \mu_{R^g}(x_i, x_j), \\ \forall x_i, x_j &\in X, \end{aligned} \quad (6)$$

and the composition $R^d \circ R^g$ of two fuzzy relations R^d and R^g is defined as

$$\begin{aligned} \mu_{R^d \circ R^g}(x_i, x_k) &= \bigvee_{x_j \in X} [\mu_{R^d}(x_i, x_j) \wedge \mu_{R^g}(x_j, x_k)], \\ \forall x_i, x_k &\in X. \end{aligned} \quad (7)$$

It should be noted that the transitive closure \bar{T} of a fuzzy tolerance relation T is a fuzzy similarity relation S .

Let α be the α -level value of the fuzzy tolerance relation T , $\alpha \in (0, 1]$. The columns and rows of the fuzzy tolerance matrix (relation) are fuzzy sets $\{A^1, \dots, A^n\}$ on X . Let $l \in \{1, \dots, n\}$, be a fuzzy set on X with $\mu_{A^l}(x_i) \in [0, 1]$, $\forall x_i \in X$, being its membership function. The α -level fuzzy set

$$A_{(\alpha)}^l = \left\{ (x_i, \mu_{A^l}(x_i)) / \mu_{A^l}(x_i) \geq \alpha, x_i \in X \right\}$$

a fuzzy α -cluster. So, $A_{(\alpha)}^l \subseteq A^l$, $\alpha \in (0, 1]$, $A^l \in \{A^1, \dots, A^n\}$, and $\mu_{A^l}(x_i)$ is the degree of membership of the element $x_i \in X$ in some fuzzy α -cluster $A_{(\alpha)}^l$, $\alpha \in (0, 1]$, $l \in \{1, \dots, n\}$. This degree of membership will be denoted by μ_{li} , for simplicity. The value of α is a tolerance threshold of elements of the fuzzy α -cluster. The membership degree of an element $x_i \in X$ in some fuzzy α -cluster $A_{(\alpha)}^l$, $\alpha \in (0, 1]$, $l \in \{1, \dots, n\}$, can be defined as a

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A_{(\alpha)}^l \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

where the α -level $A_\alpha^l = \{x_i \in X \mid \mu_{li}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$, of the fuzzy set A^l is the support of the fuzzy α -cluster A_α^l .

The value of the membership function of each element of the fuzzy α -cluster is the degree of similarity of the object to some typical object (member) of the fuzzy α -cluster.

Moreover, this membership degree defines a possibility distribution function for some fuzzy α -cluster A_α^l , $\alpha \in (0,1]$, to be denoted by $\pi_l(x_i)$.

Let $\{A_\alpha^1, \dots, A_\alpha^n\}$ be the family of fuzzy α -clusters for some α , $\alpha \in (0,1]$. The point $\tau_e^l \in A_\alpha^l$ such that

$$\tau_e^l = \arg \max_{x_i} \mu_{li}, \forall x_i \in A_\alpha^l \quad (9)$$

is called a *typical point* of the fuzzy α -cluster A_α^l , $\alpha \in (0,1]$, $l \in [1, n]$. Obviously, a fuzzy α -cluster can have several typical points, and therefore the symbol e is the index of a particular typical point.

Let $R_z^\alpha(X) = \{A_\alpha^l \mid l = \overline{1, c}, 2 \leq c \leq n\}$ be the family of fuzzy α -clusters, for some value $\alpha \in (0,1]$ of the tolerance threshold, which are generated by a fuzzy tolerance relation T on the initial set of elements $X = \{x_1, \dots, x_n\}$.

If the condition

$$\sum_{l=1}^c \mu_{li} > 0, \forall x_i \in X \quad (10)$$

is met for all A_α^l , $l = \overline{1, c}$, $c \leq n$, then this family is an *allotment* of elements of the set $X = \{x_1, \dots, x_n\}$ among the fuzzy α -clusters $\{A_\alpha^l, l = \overline{1, c}, 2 \leq c \leq n\}$, for some value $\alpha \in (0,1]$ of the tolerance threshold. It should be noted that several allotments $R_z^\alpha(X)$ can exist for any tolerance threshold. That is why the symbol z is used as the index of a particular allotment.

The allotment among the fuzzy α -clusters can be considered as a possibilistic partition and the fuzzy α -clusters meant in the sense of (8) are elements of a possibilistic partition (cf. Krishnapuram and Keller [14]). However, our next analyses will proceed using the concept of an allotment as introduced above.

A next relevant concept will now be introduced. An allotment $R_l^\alpha(X) = \{A_\alpha^l \mid l = \overline{1, n}, \alpha \in (0,1]\}$ of the set of objects among n fuzzy α -clusters for some threshold α is the initial allotment of the set $X = \{x_1, \dots, x_n\}$. In other words, if the initial data are represented by a fuzzy tolerance matrix (relation) T , then the rows and/or columns of this matrix are fuzzy sets $A^l \subseteq X$, $l = \overline{1, n}$, and the α -level fuzzy sets $A_\alpha^l \subseteq X$, $l = \overline{1, n}$, and the α -level fuzzy sets A_α^l , $l = \overline{1, n}$, $\alpha \in (0,1]$, are the fuzzy α -clusters. These fuzzy α -clusters constitute an initial allotment for some tolerance threshold $\alpha \in (0,1]$ and they can be considered as clustering components.

If some allotment $R_z^\alpha(X) = \{A_\alpha^l \mid l = \overline{1, c}, c \leq n\}$ is considered to be appropriate for the formulation of a specific problem under consideration, then this allotment is an adequate allotment. In particular, if the conditions

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

$$\text{card}(R_z^\alpha(X)) = c, \quad (11)$$

and

$$\text{card}(A_\alpha^l \cap A_\alpha^m) \leq w, \forall A_\alpha^l, A_\alpha^m, l \neq m, \alpha \in (0,1] \quad (12)$$

are met for all fuzzy α -clusters A_α^l , $l = \overline{1, c}$, of some allotment $R_z^\alpha(X) = \{A_\alpha^l \mid l = \overline{1, c}, c \leq n\}$, then this is the allotment among the particular separate fuzzy α -clusters, and $w \in \{0, \dots, n\}$ is the maximum number of elements in the intersection of different fuzzy α -clusters. If $w = 0$ in the conditions (11) and (12), then this is the allotment among fully separate fuzzy α -clusters.

An adequate allotment $R_z^\alpha(X)$, for some value $\alpha \in (0,1]$ of the tolerance threshold, is a family of α -clusters which are elements of the initial allotment $R_l^\alpha(X)$ for that value of α , and the family of fuzzy α -clusters satisfies the conditions (11) and (12). Clearly, 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 classification problem considered, and $B = \{R_z^\alpha(X)\}$ depends on the parameters of that classification problem.

On the other hand, the concept of a *principal allotment* was introduced by Viattchenin [12] and defined as follows: an allotment $R_p^\alpha(X) = \{A_\alpha^l \mid l = \overline{1, c}\}$ of the set of objects among the minimal number c , $2 \leq c \leq n$, of fully separate fuzzy α -clusters, for some tolerance threshold $\alpha \in (0,1]$, is the principal allotment of the set $X = \{x_1, \dots, x_n\}$.

The selection of the unique adequate allotment from the set $B = \{R_z^\alpha(X)\}$ of adequate allotments is made on the basis of an evaluation of allotments. The criterion employed for the evaluation of allotments is

$$F(R_z^\alpha(X), \alpha) = \sum_{l=1}^c \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c, \quad (13)$$

where c is the number of fuzzy α -clusters in the allotment $R_z^\alpha(X)$ and $n_l = \text{card}(A_\alpha^l)$, $A_\alpha^l \in R_z^\alpha(X)$, is the number of elements in the support of the fuzzy α -cluster A_α^l .

The maximum value of the criterion (13) corresponds to the best allotment of objects among c fuzzy α -clusters. So, the classification problem can be characterized formally as the determination of an optimal solution $R^*(X)$ satisfying

$$R^*(X) = \arg \max_{R_z^\alpha(X) \in B} F(R_z^\alpha(X), \alpha), \quad (14)$$

where $B = \{R_z^\alpha(X)\}$ is the set of adequate allotments corresponding to the formulation of the particular classification problem considered.

3. A Fuzzy Correlation Measure

A prototype based clustering methods can be applied if the objects are represented as points in some multidimensional space $I^m(X)$. The respective data,

composed of n objects and m attributes, are then represented as $\hat{X}_{n \times m} = [\hat{x}_i^t]$, $i = 1, \dots, n$, $t = 1, \dots, m$. Let $X = \{x_1, \dots, x_n\}$ be the set of objects. Then, the data matrix can be represented as follows:

$$\hat{X}_{n \times m} = \begin{pmatrix} \hat{x}_1^1 & \hat{x}_1^2 & \dots & \hat{x}_1^m \\ \hat{x}_2^1 & \hat{x}_2^2 & \dots & \hat{x}_2^m \\ \dots & \dots & \dots & \dots \\ \hat{x}_n^1 & \hat{x}_n^2 & \dots & \hat{x}_n^m \end{pmatrix}. \quad (15)$$

that is, be represented as $\hat{X} = (\hat{x}^1, \dots, \hat{x}^m)$ containing n -dimensional column vectors \hat{x}^t , $t = 1, \dots, m$, composed of elements of the t -th column of \hat{X} . The data can be normalized as follows [15]:

$$x_i^t = \frac{\hat{x}_i^t - \min_t \hat{x}_i^t}{\max_t \hat{x}_i^t - \min_t \hat{x}_i^t}, \quad i = 1, \dots, n, \quad t = 1, \dots, m. \quad (16)$$

so that each attribute can be interpreted as a fuzzy set x^t , $t = 1, \dots, m$, with $\mu_{x^t}(x_i)$ being its membership function.

Let x^t and x^k , $t, k \in \{1, \dots, m\}$, be two fuzzy sets with their corresponding membership functions $\mu_{x^t}(x_i)$ and $\mu_{x^k}(x_i)$, respectively.

A *fuzzy correlation measure* is defined by Chaudhuri and Bhattacharya [16] as follows:

$$r(x^t, x^k) = 1 - \left(\sum_{i=1}^n \left| \frac{\mu_{x^t}(x_i)}{\left(\sum_{i=1}^n \mu_{x^t}^\lambda(x_i) \right)^{\frac{1}{\lambda}}} - \frac{\mu_{x^k}(x_i)}{\left(\sum_{i=1}^n \mu_{x^k}^\lambda(x_i) \right)^{\frac{1}{\lambda}}} \right|^\lambda \right)^{\frac{1}{\lambda}}, \quad (17)$$

where $0 < \lambda < \infty$ is a parameter. As noted by Chaudhuri and Bhattacharya [16], the computational complexity clearly increases with the increase of λ .

The matrix of fuzzy correlation coefficients for the data can be normalized as follows [similarly as in (16)]:

$$\tilde{r}(x^t, x^k) = \frac{\left(r(x^t, x^k) - \min_{t,k} r(x^t, x^k) \right)}{\max_{t,k} r(x^t, x^k) - \min_{t,k} r(x^t, x^k)}. \quad (18)$$

so that the matrix of fuzzy correlation coefficients after normalization can be viewed as the matrix of a fuzzy tolerance relation.

4. Description of the New Algorithm

The idea of the proposed new algorithm is that attributes can be classified and a typical point of each fuzzy α -cluster can be considered to constitute an informative attribute. The proposed algorithm finds an unknown number c of disjoint fuzzy α -clusters and assigns each attribute to one of the clusters. The attributes in each cluster should have a high correlation with each other while being poorly correlated with attributes in other clusters. This method uses

the fuzzy correlation (17) as the similarity measure. The proposed algorithm can be considered as some modification of Viattchenin's [11] [12] D-PAFC-TC-algorithm used for informative features selection, and will therefore be called the D-PAFC-TC-FS.

The new algorithm is basically a classification procedure which involves 10 steps:

1. Construct the matrix of the fuzzy tolerance relation $T_{m \times m} = [\mu_T(x_t, x_k)]$, $t, k = 1, \dots, m$, by normalizing the initial data $\hat{X}_{n \times m} = [\hat{x}_i^t]$, $i = 1, \dots, n$, $t = 1, \dots, m$, according to (16) and (17);
2. Construct the transitive closure \tilde{T} of the fuzzy tolerance relation T due to (4) – (7);
3. Construct the ordered sequence of α -levels, for $0 < \alpha_0 < \alpha_1 < \dots < \alpha_l < \dots < \alpha_z \leq 1$, by the decomposition of the transitive closure \tilde{T} of the fuzzy tolerance relation T ;
4. Construct the fuzzy relation $\tilde{T}_{(\alpha_1)}$ for the consecutive values of α_1 , $\alpha_1 \in [0, 1]$;
5. Construct the initial allotment $R_i^{\alpha_1}(X) = \{A_{(\alpha_1)}^i\}$ for the fuzzy relation $\tilde{T}_{(\alpha_1)}$; construct the allotments which satisfy the conditions (11) and (12), for $w = 0$;
6. Construct the class of possible solutions of the classification problem $B(\alpha_1) = \{R_z^{\alpha_1}(X)\}$ and calculate the value of criterion (13) for each allotment $R_z^{\alpha_1}(X) \in B(\alpha_1)$;
7. Check:
 - if for some unique allotment $R_z^{\alpha_1}(X) \in B(\alpha_1)$ the condition (14) is met
 - then the allotment is the result of classification $R_p^{\alpha_1}(X)$ for the value α_1 sought, and STOP
 - else select the subset of allotments $B'(\alpha_1) \subseteq B(\alpha_1)$ which satisfy the condition (14) and go to step 8;
8. Perform the following operations for each allotment $R_z^{\alpha_1}(X) \in B'(\alpha_1)$:
 - 8.1 Set $l := 1$;
 - 8.2 Find the support $Supp(A_{(\alpha_1)}^l) = A_{\alpha_1}^l$ of the fuzzy α -cluster $A_{\alpha_1}^l \in R_z^{\alpha_1}(X)$ and construct the matrix of attributes $X_{n_l \times m} = [x_i^t]$, $x^t \in A_{\alpha_1}^l$, $i = 1, \dots, n$, for $A_{\alpha_1}^l$, where $n_l = card(A_{\alpha_1}^l)$;
 - 8.3 Calculate a prototype $\tau^l = \{\bar{x}_1, \dots, \bar{x}_n\}$ of class $A_{\alpha_1}^l$ according to the formula:

$$\bar{x}_i = \frac{1}{n_l} \sum_{x^t \in A_{\alpha_1}^l} x_i^t, \quad i = 1, \dots, n;$$
 - 8.4 Calculate the fuzzy correlation $r(\tau^l, \bar{\tau}^l)$ between the typical point τ^l of the fuzzy α -cluster $A_{(\alpha_1)}^l$ and its prototype $\bar{\tau}^l$;
 - 8.5 Check:
 - if not all fuzzy α -clusters $A_{(\alpha_1)}^l \in R_{c(z)}^{\alpha_1}(X)$ have been tested
 - then set $l := l + 1$ and go to step 8.2
 - else go to step 9
 9. Compare the fuzzy α -clusters $A_{(\alpha_1)}^l$ which are elements of different allotments $R_z^{\alpha_1}(X) \in B'(\alpha_1)$, and take the allotment $R_z^{\alpha_1}(X) \in B'(\alpha_1)$ in which the fuzzy correlation $r(\tau^l, \bar{\tau}^l)$ is minimal for all fuzzy α -clusters $A_{(\alpha_1)}^l$ obtained as a results of the classification $R_p^{\alpha_1}(X)$;

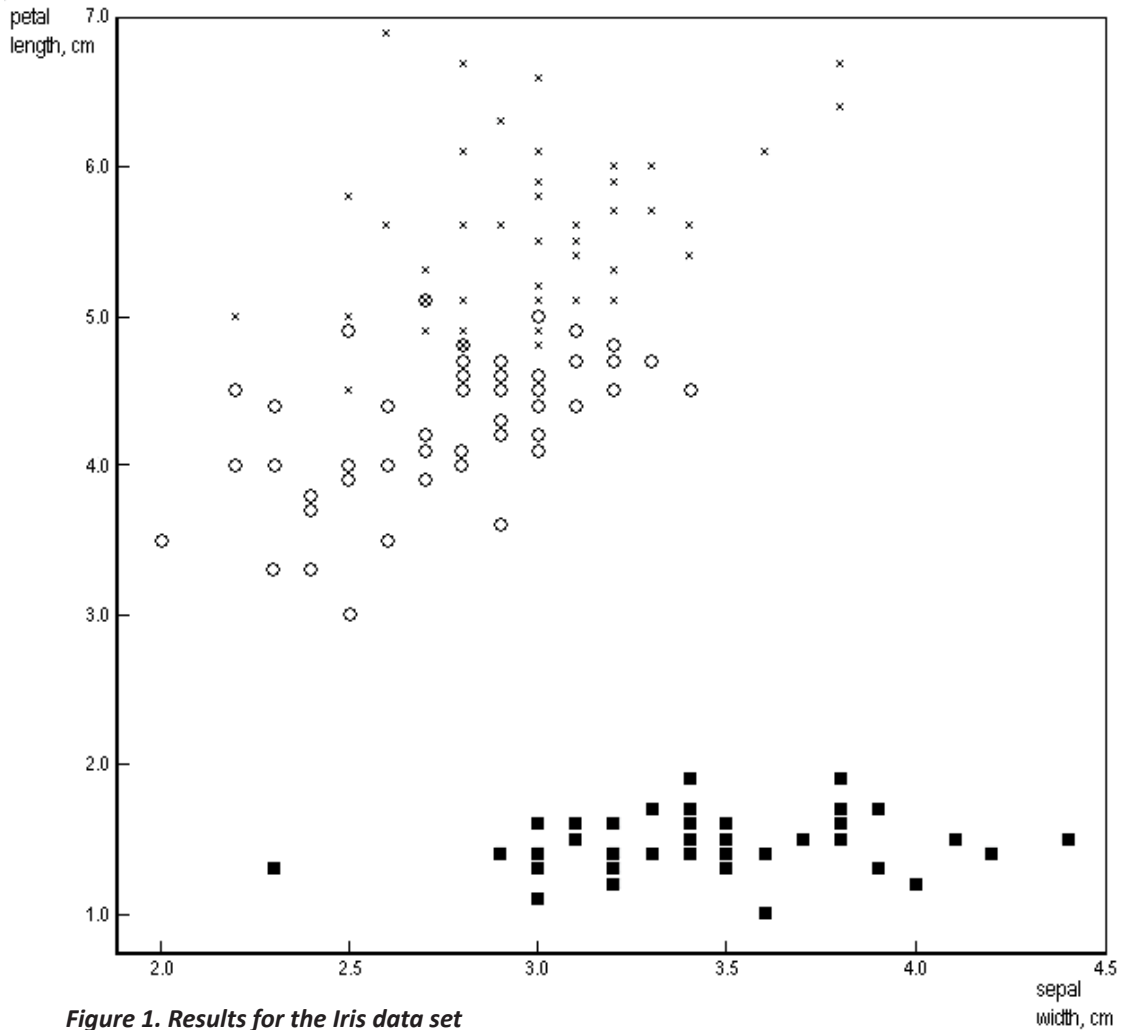


Figure 1. Results for the Iris data set

10. Select as the most informative attributes the typical points of fuzzy α -clusters of the principal allotment $R_p^{\alpha_1}(X)$ obtained.

The results of the classification sought are the typical points of the obtained principal allotment $R_p^{\alpha_1}(X) = \{A'_{(c_1)} | l = \overline{1, c}\}$ among fully separate fuzzy α -clusters and the value of the tolerance threshold $\alpha_1 \in (0, 1]$.

5. An Illustrative Example

The application of the D-PAFC-TC-FS-algorithm to feature selection can be illustrated on the well-known Iris data benchmark which was presumably first given by Anderson [17]. The Iris data set concerns the different types of Iris flowers and consists of values of the sepal length, sepal width, petal length and petal width for 150 Iris varieties. The problem is to classify the plants into three subspecies on the basis of this information. Let us consider the problem of most informative feature selection in the setting of this data set.

The Iris data set forms the matrix of attributes $X_{4 \times 150} = [\hat{x}_i^t]$, $i = 1, \dots, 150$, $t = 1, \dots, 4$, in which the sepal length vector is denoted by \hat{x}^1 , the sepal width vector is denoted by \hat{x}^2 , the petal length vector is denoted by \hat{x}^3 , and the petal width vector is denoted by \hat{x}^4 .

The D-PAFC-TC-FS-algorithm was applied directly to the normalized matrix of attributes for different values of the parameter λ , $0 < \lambda < \infty$. The principal allotment among two fuzzy α -clusters was ob-

tained in all cases. For example, the results of using the D-AFC-TC-FS-algorithm for $\lambda = 1$ are presented in Table 1. The corresponding principal allotment $R_p^{0.6692}(X)$ was obtained for the value of the tolerance threshold equal $\alpha = 0.6692$.

By executing the D-PAFC-TC-FS-algorithm for $\lambda = 2$ we also obtain two fuzzy α -clusters in the principal allotment $R_p^{0.5583}(X)$. The matrix of the principal allotment $R_p^{0.5583}(X)$ is presented in Table 2.

For $\lambda = 3$ we obtain the principal allotment $R_p^{0.4917}(X)$ among two fuzzy α -clusters. The corresponding matrix is presented in Table 3.

The second feature \hat{x}^2 is the typical point of the first fuzzy α -cluster and the third feature \hat{x}^3 is the typical point of the second fuzzy α -cluster in each case. Obviously, the features \hat{x}^2 and \hat{x}^3 can be selected as the most informative features. So, the two-dimensional projection of the Iris data can be constructed as presented in Figure 1.

Two well separated classes can be distinguished, and then visualized. The first class corresponds to the Iris Setosa. The second class corresponds to the Iris Versicolor and Iris Virginica. The objects known to be the Iris Setosa are represented by \blacksquare in Figure 1, those known to be the Iris Versicolor are represented by \circ , and those known to be the Iris Virginica are represented by \times .

It is worth noticing that the result is similar to that obtained by using the conventional method of the prin-

Table 1.

The memberships values obtained by the D-PAFC-TC-FS-algorithm for $\lambda=1$

Class	Features			
	\hat{x}^1	\hat{x}^2	\hat{x}^3	\hat{x}^4
1	0.0000	1.0000	0.0000	0.0000
2	0.6692	0.0000	1.0000	0.8347

Table 2

The memberships values obtained by the D-PAFC-TC-FS-algorithm for $\lambda=2$

Class	Features			
	\hat{x}^1	\hat{x}^2	\hat{x}^3	\hat{x}^4
1	0.0000	1.0000	0.0000	0.0000
2	0.5583	0.0000	1.0000	0.7635

Table 3

The memberships values obtained by the D-PAFC-TC-FS-algorithm for $\lambda=3$

Class	Features			
	\hat{x}^1	\hat{x}^2	\hat{x}^3	\hat{x}^4
1	0.0000	1.0000	0.0000	0.0000
2	0.4917	0.0000	1.0000	0.7161

Table 4

Factor loads in the principal component analysis

Features	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

principal component analysis (cf. Sato-Ilic and Jain [18]). An interpretation of the obtained principal components can be made on the basis of factor loading. The factor loading is defined as a correlation coefficient between the v -th principal component z_v and the t -th attribute \hat{x}^t , $t = 1, \dots, m$, as follows [18]:

$$f(z_v, \hat{x}^t) = \frac{\text{cov}\{z_v, \hat{x}^t\}}{\sqrt{V\{z_v\}V\{\hat{x}^t\}}}, \quad (19)$$

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

In Table 4 the values of factor loadings (19) are shown which can represent a relationship between each principal component and each attribute.

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

In Table 4 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 of sepal width.

From the comparison between the results obtained and shown in Tables 1, 2, 3 and 4, we can see similar outcomes. In particular, in Table 4, values of the membership function of the first fuzzy α -cluster of the principal allotment in each case can be interpreted as the normalized values of factor loadings $f(z_t, \hat{x}^t)$, $t = 1, \dots, 4$, and values of the membership function of the second fuzzy α -cluster of the principal allotment can be considered as the normalized values of factor loadings $f(z_1, \hat{x}^t)$, $t = 1, \dots, 4$.

6. Concluding Remarks

First, from a more general point of view, one should mention that the concepts of a fuzzy α -cluster and allotment have quite a clear epistemological motivation. That is why the results of application of the possibilistic clustering method based on the concept of allotment can be very well interpreted.

The D-PAFC-TC-FS-algorithm of possibilistic clustering is proposed in this paper. The determination of an unknown number of the most informative features is the main feature of the proposed algorithm. The D-PAFC-TC-FS-algorithm can be considered as a version of the method of extremal grouping of features. The algorithm is based on the concept of a principal allotment among the fuzzy α -clusters, and an unknown minimal number of compact and well-separated fuzzy α -clusters is the result of classification. The typical points of the fuzzy α -clusters can be selected as the most informative features. Moreover, the D-PAFC-TC-FS-algorithm does not depend on parameters and can be applied directly to the data given as a matrix of attribute values. The results of application of the algorithm proposed to the Iris data show that the D-PAFC-TC-FS-algorithm is an effective and efficient numerical procedure for solving the informative feature selection problem.

As for future extensions, for instance, the proposed D-PAFC-TC-FS-algorithm can be extended by including different fuzzy correlation measures exemplified by those proposed by Murthy, Pal and Dutta Majumder [19], and Chiang and Lin [20].

Acknowledgements

The research has been partially supported by, first, the National Centre of Science under Grant No. UMO-2011/01/B/ST6/06908 (J. Kacprzyk and J.W. Owsinski), and second, the Mianowski Fund and the Ministry of Foreign Affairs of the Republic of Po-

land (D. Viattchenin). The authors are grateful to Mr. Aliaksandr Damaratski for developing software used for numerical experiments, and to anonymous referees for their valuable comments.

AUTHORS

Janusz Kacprzyk*, **Jan W. Owsinski** – ¹Systems Research Institute, Polish Academy of Sciences
6 Newelska St., 01-447 Warsaw, Poland.

E-mails: {kacprzyk, owsinski}@ibspan.waw.pl

Dimitri A. Viattchenin – United Institute of Informatics Problems, National Academy of Sciences of Belarus
6 Surganov St., 220012 Minsk, Belarus.

E-mail: viattchenin@mail.ru

*Corresponding author

REFERENCES

- [1] Blum A. , Langley P., "Selection of relevant features and examples in machine learning", *Artificial Intelligence*, vol. 97, no.1-2, 1997, 245-271. DOI: [http://dx.doi.org/10.1016/S0004-3702\(97\)00063-5](http://dx.doi.org/10.1016/S0004-3702(97)00063-5)
- [2] Kohavi R. , John G., "Wrappers for feature subset selection", *Artificial Intelligence*, vol. 97, no. 1-2, 1997, 273-324. DOI: [http://dx.doi.org/10.1016/S0004-3702\(97\)00043-X](http://dx.doi.org/10.1016/S0004-3702(97)00043-X)
- [3] Ghazavi S.N., Liao T.W., "Medical data mining by fuzzy modeling with selected features", *Artificial Intelligence in Medicine*, vol. 43, no. 3, 2008, 195-206. DOI: <http://dx.doi.org/10.1016/j.art-med.2008.04.004>
- [4] Draminski M., Kierczak M., Nowak-Brzezinska A., Koronacki J., and Komorowski J., "The Monte Carlo feature selection and interdependency discovery is unbiased", *Control and Cybernetics*, vol. 40, no. 2, 2011, 199-211.
- [5] Kong Y.-Q., Wang S.-T., "Feature selection and semi-supervised fuzzy clustering", *Fuzzy Information and Engineering*, vol. 1, no. 2, 2009, 179-190.
- [6] Chitsaz E., Taheri M., Katebi S.D., "A fuzzy approach to clustering and selecting features for classification of gene expression data". In: *Proc. World Congress of Engineering (WCE 2008)*, 2008, 1650-1655.
- [7] Chitsaz E., Taheri M., Katebi S.D., Jahromi M.Z., "An improved fuzzy feature clustering and selection based on chi-squared-test". In: *Proc. Int. Multiconference of Engineers and Computer Scientists (IMECS 2009)*, 2009, 35-40.
- [8] Bezdek J.C., Keller J.M., Krishnapuram R., Pal N.R., *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer Science, New York, 2005. DOI: <http://dx.doi.org/10.1007/b106267>.
- [9] Mandel I.D., *Clustering Analysis*, Finansy i Statistika, Moscow, 1988. (in Russian)
- [10] Viattchenin D.A., "A new heuristic algorithm of fuzzy clustering", *Control and Cybernetics*, vol. 33, no. 2, 2004, 323-340.
- [11] Viattchenin D.A., "A direct algorithm of possibilistic clustering with partial supervision", *Journal of Automation, Mobile Robotics and Intelligent Systems*, vol. 1, no. 3, 2007, 29-38.
- [12] Viattchenin D.A., "An algorithm for detecting the principal allotment among fuzzy clusters and its application as a technique of reduction of analyzed features space dimensionality", *J. Information and Organizational Sciences*, vol. 33, no. 1, 2009, 205-217.
- [13] Viattchenin D.A., "Direct algorithms of fuzzy clustering based on the transitive closure operation and their application to outliers detection", *Artificial Intelligence*, no. 3, 2007, 205-216. (in Russian)
- [14] Krishnapuram R., Keller J.M., "A possibilistic approach to clustering", *IEEE Trans. on Fuzzy Systems*, vol. 1, no. 2, 1993, 98-110. DOI: <http://dx.doi.org/10.1109/91.227387>
- [15] Walesiak M., *A generalized distance measure in statistical multivariate analysis.*, Pub. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław, 2002 (in Polish).
- [16] Chaudhuri B.B., Bhattacharya A., "On correlation between two fuzzy sets", *Fuzzy Sets and Systems*, vol. 118, no. 3, 2001, 447-456.
- [17] Anderson E., "The irises of the Gaspé Peninsula", *Bulletin of the American Iris Society*, vol. 59, no. 1, 1935, 2-5.
- [18] Sato-Ilic M., Jain L.C., *Innovations in Fuzzy Clustering: Theory and Applications*, Springer-Verlag, Heidelberg, 2006.
- [19] Murthy C.A., Pal S.K., Dutta Majumder D., "Correlation between two fuzzy membership functions", *Fuzzy Sets and Systems*, vol. 17, no. 1, 1985, 23-38. DOI: [http://dx.doi.org/10.1016/0165-0114\(85\)90004-1](http://dx.doi.org/10.1016/0165-0114(85)90004-1)
- [20] Chiang D.-A., Lin N.P., "Correlation of fuzzy sets", *Fuzzy Sets and Systems*, vol. 102, no. 2, 1999, 221-226. DOI: [http://dx.doi.org/10.1016/S0165-0114\(97\)00127-9](http://dx.doi.org/10.1016/S0165-0114(97)00127-9)