www.czasopisma.pan.pl PAN POLSKA AKADEMIA NAUK www.journals.pan.pl

INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2017, VOL. 63, NO. 4, PP.341-346
Manuscript received January 13, 2017; revised September, 2017. DOI: 10.1515/eletel-2017-0046

A Method for Estimating the Least Number of Objects in Fuzzy Clusters

Dmitri A. Viattchenin and Aliaksandr Yaroma

***Abstract*—The theoretical note deals with the problem of estimation of the value of the least number of objects in fuzzy clusters for following detection of the optimal number of objects in fuzzy clusters through heuristic possibilistic clustering. A technique for detecting the optimal maximal number of elements in the a priori unknown number of fuzzy clusters of the sought clustering structure is reminded and a procedure for finding the initial minimal value of the number of objects in fuzzy clusters is proposed. Numerical examples are considered and conclusions are formulated.**

***Keywords*—possibilistic clustering, fuzzy cluster, allotment, cluster size**

I. Introduction

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 pair wise disjoint, all non empty and reproduce X via union. Clustering methods have been applied effectively in signal processing, image recognition and telecommunications.

Since the fundamental Zadeh's paper [1] was published, fuzzy sets theory has been applied to many areas such as learning, decision-making and classification. Heuristic methods of fuzzy clustering, hierarchical methods of fuzzy clustering and optimization methods of fuzzy clustering were proposed by different researchers. Moreover, a possibilistic approach to clustering was proposed by Krishnapuram and Keller in [2] and developed by other researchers. A concept of possibilistic partition is a basis of possibilistic clustering methods and the membership values can be interpreted as the values of typicality degree. Fuzzy and possibilistic clustering methods are considered at length, for instance, in [3-5].

The most common and widespread approach to fuzzy clustering is the optimization approach. Moreover, major possibilistic clustering methods are also objective function-based clustering algorithms. However, heuristic algorithms of fuzzy clustering are simple and very effective in many cases, because heuristic algorithms display high level of essential clarity and low level of complexity. Some heuristic clustering procedures are based on the definition of a cluster concept and the purpose of these algorithms is cluster detection conform to a given definition. Such algorithms are called algorithms of direct classification or direct clustering algorithms. Thus, a heuristic approach to possibilistic clustering in which the sought clustering structure of the set of objects is based directly on the formal definition of fuzzy α-cluster and possibilistic memberships are determined directly from the values of pairwise similarity of objects was proposed in [6] and developed in other publications.

Different classification techniques were proposed in the framework of the heuristic approach to possibilistic clustering. In particular, a technique for detecting the optimal maximal number of elements in the a priori unknown number of fuzzy clusters of the sought clustering structure is proposed in [7]. The technique is based on the direct relational D-AFC(u)-algorithm of possibilistic clustering [8] and corresponding cluster validity measures. However, a problem of finding the initial value of the least number of objects in fuzzy clusters for using the technique is not solved in [7]. Thus, the main goal of the presented paper is solving the problem of estimation of the value of the least number of objects in fuzzy clusters for detection of the optimal number of objects in fuzzy clusters through heuristic possibilistic clustering.

So, the content of this paper is as follows: in the second section basic definitions of the heuristic approach to possibilistic clustering are reminded and corresponding algorithms are enumerated, in the third section the technique for detecting the optimal maximal number of elements in the a priori unknown number of fuzzy clusters of the sought clustering structure is presented and a procedure for finding the initial value of the least number of objects in fuzzy clusters is proposed, in the fourth section illustrative examples of application of the proposed method to artificial data sets are considered and in fifth section some preliminary conclusions are made.

II. A Heuristic Approach to Possibilistic Clustering

Basic concepts of the heuristic approach to possibilistic clustering are considered in the first subsection. The second subsection includes a brief review of corresponding algorithms. Notes on the data pre-processing are given in the third subsection of the section.

A. Basic Definitions

Let us remind basic concepts of the heuristic approach to possibilistic clustering [6]. 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 are fuzzy sets $\{A^1,...,A^n\}$. Let $\{A^1,...,A^n\}$ be fuzzy sets on X, which are generated by a fuzzy tolerance T. The α-level fuzzy set $A^l_{(\alpha)}=\{(x_i,\mu_{A^l}(x_i))\mid\mu_{A^l}(x_i)\geq\alpha\}$,

D. A. Viattchenin is with Laboratory of System Identification, United Institute of Informatics Problems, National Academy of Sciences of Belarus, Minsk, Belarus, (e-mail: viattchenin@mail.ru).

A. Yaroma is with Department of Software Information Technology, Faculty of Computer Systems and Networks, Belarusian State University of Informatics and Radio-Electronics, Minsk, Belarus, (e-mail: aeroma@inbox.ru).

$l \in [1,n]$ 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 α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$. The value of α is the tolerance threshold of elements of fuzzy α-clusters.

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

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_\alpha \\ 0, & \text{otherwise} \end{cases}, \tag{1}$$

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 α-cluster $A^l_{(\alpha)}$. So, condition $A^l_\alpha = Supp(A^l_{(\alpha)})$ is met for each fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$. The value of membership function of each object of fuzzy α-cluster is the degree of similarity of the object to some typical object of fuzzy α-cluster. So, the object $\tau^l_e \in A^l_\alpha$, for which

$$\tau^l_e = \arg\max_{x_i} \mu_{li}, \ \forall x_i \in A^l_\alpha, \tag{2}$$

is called a typical point of the fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$. A fuzzy α-cluster $A^l_{(\alpha)}$ can have several typical points. That is why symbol e is the index of the typical point.

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

$$\sum_{l=1}^{c} \mu_{li} > 0, \ \forall x_i \in X \tag{3}$$

is met for all fuzzy α-clusters $A^l_{(\alpha)} \in R^\alpha_{c(z)}(X)$, $l = \overline{1,c}$, $c \leq n$, then the family is the allotment of elements of the set $X = \{x_1,\ldots,x_n\}$ among fuzzy α-clusters $\{A^l_{(\alpha)}, l = \overline{1,c}, 2 \leq c \leq n\}$ for some value of the tolerance threshold α. It should be noted that several allotments $R^\alpha_{c(z)}(X)$ can exist for some tolerance threshold α. That is why symbol z is the index of an allotment.

It should be noted, that the condition (3) equal to the condition of possibilistic partition [2]. So, the allotment of elements of the data set among fuzzy α-clusters is a particular case of the possibilistic partition.

Allotment $R^\alpha_I(X) = \{A^l_{(\alpha)} \mid l = \overline{1,n}, \alpha \in (0,1]\}$ of the set of objects among n fuzzy α-clusters for some tolerance threshold $\alpha \in (0,1]$ is the initial allotment of the set $X = \{x_1,\ldots,x_n\}$. Simply, if initial data are represented by a matrix of some fuzzy tolerance T then lines or columns of the matrix are fuzzy sets $A^l \subseteq X$, $l = \overline{1,n}$ and α-level fuzzy sets $A^l_{(\alpha)}$, $l = \overline{1,c}$, $\alpha \in (0,1]$ are fuzzy α-clusters. These fuzzy α-clusters can be considered as clustering components.

If some allotment $R^\alpha_{c(z)}(X) = \left\{A^l_{(\alpha)} \mid l = \overline{1,c}, c \leq n, \alpha \in (0,1]\right\}$ corresponds to the formulation of a concrete problem, then this allotment is an adequate allotment. In particular, if a condition

$$\begin{aligned} &\sum_{l=1}^{c} card(A^l_\alpha) \geq card(X), \quad \forall A^l_{(\alpha)} \in R^\alpha_{c(z)}(X), \\ &\alpha \in (0,1], \quad card(R^\alpha_{c(z)}(X)) = c \end{aligned}, \tag{4}$$

and a condition

$$card(A^l_\alpha \cap A^m_\alpha) \leq w, \ \forall A^l_{(\alpha)}, A^m_{(\alpha)}, \ l \neq m, \ \alpha \in (0,1], \tag{5}$$

are met for all fuzzy clusters $A^l_{(\alpha)}$, $l = \overline{1,c}$ of some allotment $R^\alpha_{c(z)}(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, c \leq n\}$ for a value $\alpha \in (0,1]$, then the allotment is the allotment among particularly separate fuzzy clusters. The value $w \in \{0,\ldots,n\}$ is the maximum number of elements in the intersection area of different fuzzy clusters. For $w = 0$ fuzzy clusters are fully separate fuzzy clusters.

The adequate allotment $R^\alpha_{c(z)}(X)$ for some value of tolerance threshold $\alpha \in (0,1]$ is a family of fuzzy clusters which are elements of the initial allotment $R^\alpha_I(X)$ for the value of α and the family of fuzzy clusters should satisfy the conditions (4) and (5). So, the construction of adequate allotments $R^\alpha_{c(z)}(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, c \leq n\}$ for every α is a trivial problem of combinatorics.

Allotment $R^\alpha_P(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}\}$ of the set of objects among the minimal number c, $2 \leq c \leq n$ of fully separate fuzzy clusters for some tolerance threshold $\alpha \in (0,1]$ is the principal allotment of the set $X = \{x_1,\ldots,x_n\}$.

Several adequate allotments can exist. Thus, the problem consists in the selection of the unique adequate allotment $R^*_c(X)$ from the set B of adequate allotments, $B = \{R^\alpha_{c(z)}(X)\}$, which is the class of possible solutions of the concrete classification problem. The selection of the unique adequate allotment $R^*_c(X)$ from the set $B = \{R^\alpha_{c(z)}(X)\}$ of adequate allotments must be made on the basis of evaluation of allotments. In particular, the criterion

$$F(R^\alpha_{c(z)}(X),\alpha) = \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c, \tag{6}$$

where c is the number of fuzzy α-clusters in the allotment $R^\alpha_{c(z)}(X)$ and $n_l = card(A^l_\alpha)$, $A^l_{(\alpha)} \in R^\alpha_{c(z)}(X)$ is the number of elements in the support of the fuzzy α-cluster $A^l_{(\alpha)}$, can be used for evaluation of allotments. Maximum of criterion (6) corresponds to the best allotment of objects among c fuzzy α-clusters. So, the classification problem can be characterized formally as determination of the solution $R^*_c(X)$ satisfying

$$R^*_c(X) = \arg\max_{R^\alpha_{c(z)}(X) \in B} F(R^\alpha_{c(z)}(X),\alpha). \tag{7}$$

The problem of cluster analysis can be defined in general as the problem of discovering the unique allotment $R_c^*(X)$, resulting from the classification process.

B. A Brief Review of Clustering Procedures

Direct heuristic algorithms of possibilistic clustering can be divided into two types: relational versus prototype-based. A fuzzy tolerance relation T matrix is a matrix of the initial data for the direct heuristic relational algorithms of possibilistic clustering and a matrix of attributes is a matrix of the initial data for the prototype-based algorithms. In particular, the group of direct relational heuristic algorithms of possibilistic clustering includes

- the D-AFC(c)-algorithm which is based on the construction of an allotment $R_c^*(X)$ among an a priori given number c of partially separate fuzzy α-clusters [6];
- the D-PAFC-algorithm which is based on the construction of an principal allotment $R_P^*(X)$ among an unknown minimal number of at least c fully separate fuzzy α-clusters [6];
- the D-AFC-PS(c)-algorithm which is based on the construction of an allotment $R_c^*(X)$ among an a priori given number c of partially separate fuzzy α-clusters in the presence of labelled object [6];
- the D-AFC(u)-algorithm which is based on the construction of an allotment $R_c^*(X)$ among an a priori unknown number c of partially separate fuzzy α-clusters with respect to the given maximal number u of elements in every class [8].
- the D-AFC(α)-algorithm which is based on the construction of an allotment $R_c^*(X)$ among an a priori unknown number c of partially separate fuzzy α-clusters with respect to the given minimal value α of tolerance threshold [9];

Moreover, the FG-AFC-algorithm of heuristic possibilistic clustering based on fuzzy tolerance graph decomposition was proposed in [10].

On the other hand, the family of direct prototype-based heuristic algorithms of possibilistic clustering includes [6]

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

The hierarchical H-AFC-TC-algorithm which is based on the construction of a hierarchy of allotments among an a priori unknown number c of fully separate fuzzy α-clusters was also proposed in [6].

It should be noted, that these prototype-based heuristic algorithms of possibilistic clustering are based on the transitive closure of the initial fuzzy tolerance. On the other hand, a family of direct prototype-based heuristic possibilistic clustering algorithms based on a transitive approximation of a fuzzy tolerance is proposed in [11]. So, direct prototype-based heuristic possibilistic clustering algorithms which based on a transitive closure of an initial fuzzy tolerance are a special case of corresponding clustering procedures which based on a transitive approximation of a fuzzy tolerance.

C. Notes on the Data Pre-Processing

In the relational approach to clustering, the problem of the data classification is solved by expressing a relation which quantifies either similarity, or dissimilarity, between pairs of objects. So, the data matrix taken a form

$$\hat{\rho}_{n\times n} = \begin{pmatrix} \hat{\rho}_{11} & \hat{\rho}_{12} & \dots & \hat{\rho}_{1n} \\ \hat{\rho}_{21} & \hat{\rho}_{22} & \dots & \hat{\rho}_{2n} \\ \dots & \dots & \dots & \dots \\ \hat{\rho}_{n1} & \hat{\rho}_{n2} & \dots & \hat{\rho}_{nn} \end{pmatrix}, \tag{8}$$

where a general notation $\hat{\rho}_{ij}$ used for designation of pair wise dissimilarities $d(x_i, x_j)$ or the similarity coefficients $r(x_i, x_j)$. In general, the values $\hat{\rho}_{ij}$ are not normalized. The relational data can be normalized as follows:

$$\rho_{ij} = \frac{\left(\hat{\rho}_{ij} - \min\limits_{i,j} \hat{\rho}_{ij}\right)}{\max\limits_{i,j} \hat{\rho}_{ij} - \min\limits_{i,j} \hat{\rho}_{ij}}, \tag{9}$$

where the general notation $\hat{\rho}_{ij}$ is used for designation of pair-wise dissimilarities or the similarity coefficients. If $\rho_{ii} = 0$, $\forall i$, then the relation matrix $\rho_{n\times n} = [\rho_{ij}]$ is the matrix of fuzzy intolerance $I = [\mu_I(x_i, x_j)]$, $i, j = 1, \dots, n$.

On the other hand, the object data clustering methods can be applied if the objects are represented as points in some multidimensional space $I^{m_1}(X)$. In other words, the data which is composed of n objects and m_1 attributes is denoted as $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1}]$, $i = 1, \dots, n$, $t_1 = 1, \dots, m_1$ and the data are called sometimes the two-way data [12].

Let $X = \{x_1, ..., x_n\}$ is the set of objects. So, the two-way data matrix can be represented as follows:

$$\hat{X}_{n\times m_1} = \begin{pmatrix} \hat{x}_1^1 & \hat{x}_1^2 & \dots & \hat{x}_1^{m_1} \\ \hat{x}_2^2 & \hat{x}_2^2 & \dots & \hat{x}_2^{m_1} \\ \dots & \dots & \dots & \dots \\ \hat{x}_n^1 & \hat{x}_n^2 & \dots & \hat{x}_n^{m_1} \end{pmatrix}. \tag{10}$$

Thus, the two-way data matrix can be represented as $\hat{X} = (\hat{x}^1, \dots, \hat{x}^{m_1})$ using n-dimensional column vectors $\hat{x}^{t_1}$, $t_1 = 1, \dots, m_1$, composed of the elements of the t_1-th column of $\hat{X}$.

The matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \dots, n$ is the matrix of initial data for the relational heuristic algorithms of possibilistic clustering. However, the initial data can be presented as a matrix of attributes $\hat{X}_{n\times m_1} = [\hat{x}_i^{t_1}]$, $i = 1, \dots, n$, $t_1 = 1, \dots, m_1$, where the value $\hat{x}_i^{t_1}$ is the value of the t_1-th attribute for i-th object. Thus, the proposed approach to clustering can be used with the two-way data (10), by choosing

a suitable metric to measure similarity. However, the initial data should be normalized.

In the first place, the two-way data can be normalized as follows:

$$x_i^{t_1} = \frac{\hat{x}_i^{t_1}}{\max\limits_i \hat{x}_i^{t_1}}. \quad (11)$$

The data normalization method (11) is appropriate in the case of non-negative values $\hat{x}_i^{t_1}$ in the two-way data matrix.

In the second place, the two-way data can be normalized using a formula

$$x_i^{t_1} = \frac{\hat{x}_i^{t_1} - \min\limits_i \hat{x}_i^{t_1}}{\max\limits_i \hat{x}_i^{t_1} - \min\limits_i \hat{x}_i^{t_1}}. \quad (12)$$

So, each object can be considered as a fuzzy set x_i, $i = 1,\ldots,n$ and $x_i^{t_1} = \mu_{x_i}(x^{t_1}) \in [0,1]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$ are their membership functions.

Some other methods for the two-way data normalization are described in bibliographical sources. Different methods for the data normalization are considered, for example, by Walesiak [13].

The matrix of coefficients of pair wise dissimilarity between objects $I = [\mu_I(x_i, x_j)]$, $i, j = 1,\ldots,n$ can be obtained after application of some distance function to the matrix of normalized data $X_{n\times m_1} = [\mu_{x_i}(x^{t_1})]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$. The most widely used distances for fuzzy sets x_i, x_j, $i, j = 1,\ldots,n$ in $X = \{x_1,...,x_n\}$ are considered by Kaufmann in [14] and these distances can be described as follows.

- The normalized Hamming distance:

$$l(x_i, x_j) = \frac{1}{m_1}\sum_{t_1=1}^{m_1} |\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1})|,\ i, j = 1,\ldots,n. \quad (13)$$

- The normalized Euclidean distance:

$$e(x_i, x_j) = \sqrt{\frac{1}{m_1}\sum_{t_1=1}^{m_1}\left(\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1})\right)^2},\ i, j = 1,\ldots,n. \quad (14)$$

- The squared normalized Euclidean distance:

$$\varepsilon(x_i, x_j) = \frac{1}{m_1}\sum_{t_1=1}^{m_1}\left(\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1})\right)^2,\ i, j = 1,\ldots,n. \quad (15)$$

The matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1,\ldots,n$ can be obtained after application of complement operation

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j),\ i, j = 1,\ldots,n. \quad (16)$$

to the matrix of dissimilarity coefficients $I = [\mu_I(x_i, x_j)]$, $i, j = 1,\ldots,n$ obtained from previous operations.

III. An Outline of the Approach

The first subsection includes the detail consideration of the technique for detecting the optimal maximal number of elements in fuzzy α-clusters. A procedure for finding the initial value of the least number of objects in fuzzy α-clusters is proposed in the second subsection of the section.

A. The Technique for Detecting the Optimal Number of Objects in Fuzzy α-Clusters

Let us remind the essence of the D-AFC(u)-algorithm which was proposed in [8]. An analyst can determine the maximal number u of elements in a fuzzy α-cluster. If $1 \le u < n$ is a maximal number of elements in a fuzzy α-cluster, then $1 \le n_l \le u$, $\forall l = \overline{1,c}$, where $n_l = card(A_\alpha^l)$, $A_\alpha^l = Supp\left(A_{(\alpha)}^l\right)$ for each fuzzy α-cluster $A_{(\alpha)}^l$, $l = \overline{1,c}$, $\alpha \in (0,1]$. So, parameter u can be considered as the parameter that controls cluster sizes and the classification problem can be formulated as follows: detection of an unknown number c of partially separated fuzzy α-clusters with given maximal number of elements $1 \le u < n$ in every class can be considered as the aim of classification. For lack of space, a plan of the D-AFC(u)-algorithm is omitted here.

The D-AFC(u)-algorithm can be considered as an appropriate tool for detecting the optimal number of elements in fuzzy α-clusters of the constructed allotment. For the purpose, validity measures should be used because the number of fuzzy α-clusters in the sought allotment depends on the number of elements in each fuzzy α-cluster. In particular, for the D-AFC(c)-algorithm of possibilistic clustering next validity measure were proposed in [6]:

- the linear measure of fuzziness of the allotment;
- the quadratic measure of fuzziness of the allotment;
- the measure of separation and compactness of the allotment.

Using the linear measure of fuzziness of the allotment or the quadratic measure of fuzziness of the allotment, the optimal number c of fuzzy α-clusters can be obtained by maximizing the index value. Otherwise, optimum value of c is obtaining by minimizing the measure of separation and compactness of the allotment.

Let $V(R_c^{*(k)}(X); c)$ be a general notation for validity measures. The value u can vary in the interval $[u_{\min}, u_{\max}]$. Optimal number of elements in fuzzy α-clusters depends on next criteria [7]:

- fuzzy α-clusters should be as possible as more separated in the constructed allotment;
- fuzzy α-clusters must be homogeneous, that is the number of elements in fuzzy α-clusters should be approximately equal, as possible.

So, the proposed in [7] technique for detecting the optimal maximal number of elements in fuzzy α-clusters is a five-step procedure as given below:

1. Set $k := 1$ and $u_k := u_{\min}$;
2. The D-AFC(u)-algorithm should be applied to the matrix of tolerance coefficients $T = [\mu_T(x_i, x_j)]$, $i, j = 1,\ldots,n$ for the current value u_k and the corresponding allotment $R_c^{*(k)}(X)$ will be constructed;
3. Calculate the value of some validity measure $V(R_c^{*(k)}(X); c)$;
4. The following condition is checked:
   if the condition $u_k = u_{\min}$ is met
   then set $k := k+1$, $u_{k+1} := u_k + 1$ and go to step 2
   else go to step 5;

5. The following condition is checked:
if the condition $\left|V(R_c^{*(k+1)}(X);c)-V(R_c^{*(k)}(X);c)\right|=0$ is met
then the value u_k is the optimal number of elements in each fuzzy α-cluster $A_{(\alpha)}^l$, $l=1,\ldots,c$ of the constructed allotment $R_c^{*(k)}(X)$ and stop
else set $k:=k+1$, $u_{k+1}:=u_k+1$ and go to step 2;

So, the value $u_{\min}$ should be pre-determined for using the technique. This value can be defined by analyst. However, the value $u_{\min}$ can be detected automatically. The corresponding procedure is proposed in the next subsection.

B. A Procedure for Finding the Initial Minimal Number of Objects in Fuzzy α-Clusters

The initial minimal number $u_{\min}$ of objects in fuzzy α-clusters of the sought allotment can be determined as a cardinality of the support of the smallest fuzzy α-cluster which is an element of the principal allotment, $R_P^\alpha(X)$. So, a procedure for detecting the value $u_{\min}$ can be described in the following way:

1. Construct the principal allotment $R_P^\alpha(X)=\{A_{(\alpha)}^l \mid l=\overline{1,c}\}$ of the set $X=\{x_1,...,x_n\}$;
2. Calculate the value $n_l=card(A_\alpha^l)$ for each $A_{(\alpha)}^l\in R_P^\alpha(X)$, $l=1,\ldots,c$;
3. Set $u_{\min}:=\min\limits_l n_l$ and stop.

If the initial data represented by the matrix of fuzzy tolerance relation $T=[\mu_T(x_i,x_j)]$, $i,j=1,\ldots,n$, then the relational D-PAFC-algorithm should be used in the first step of the procedure. Otherwise, if the initial data are presented as a matrix of attributes $\hat{X}_{n\times m_1}=[\hat{x}_i^{t_1}]$, $i=1,\ldots,n$, $t_1=1,\ldots,m_1$, then the prototype-based D-PAFC-TC-algorithm should be used in the first step of the procedure for constructing the principal allotment $R_P^\alpha(X)$.

An application of the proposed technique to the classification problem will be illustrated on the well-known benchmark in the next section.

IV. An Illustrative Example

Let us consider an example of application of the proposed technique to the data processing. For the purpose, the Sneath and Sokal's two-dimensional data set [15] was selected. The artificial data set is shown in Fig. 1.

Let us consider the result obtained from the first step of the proposed procedure. The initial data set was normalized by using the formula (11) and the squared normalized Euclidean distance (15) was selected as a parameter for the D-PAFC-TC-algorithm in the procedure.

By executing the D-PAFC-TC-algorithm, we obtain the principal allotment $R_P^*(X)$ among $c=2$ fully separated fuzzy

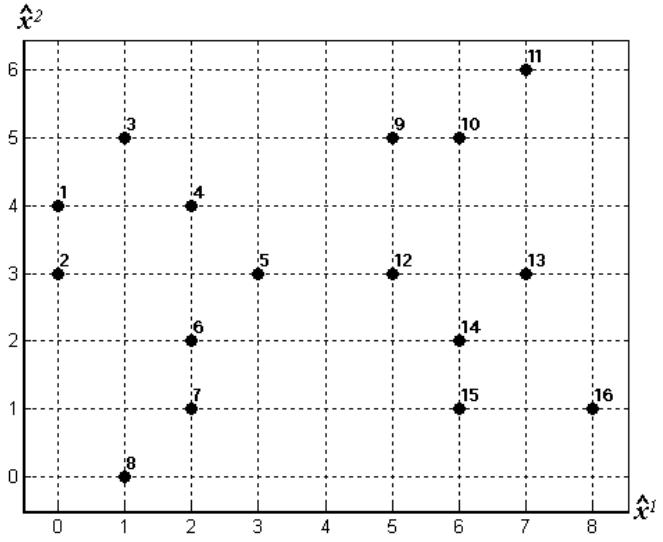

Fig. 1. Sneath and Sokal's data set.

α-clusters which corresponds to the result, obtained for the tolerance threshold $\alpha=0.96875$. The sixth object is the typical point of the first fuzzy α-cluster and the tenth object is the typical point of the second fuzzy α-cluster.

Membership functions of two fuzzy α-clusters are presented in Fig. 2, where membership values of the first class are represented by ○ and membership values of the second class are represented by ■. The obtained result was also presented in [6].

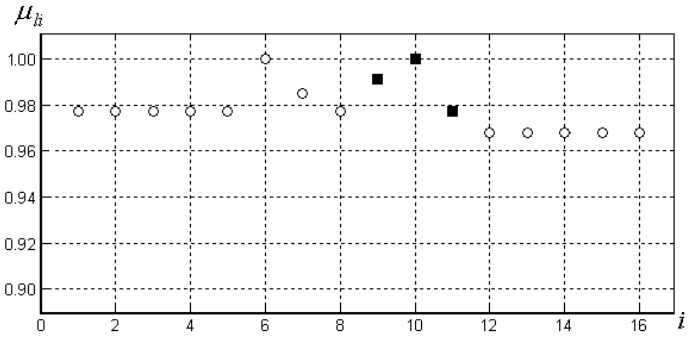

Fig. 2. Membership functions of two fuzzy clusters obtained from the D-PAFC-TC-algorithm.

So, the value of the minimal number $u_{\min}$ of objects in fuzzy α-clusters is equal 3 and it is the result obtained from the proposed procedure. That is why $u_{\min}=3$ is the input value for the technique for detecting the optimal maximal number of elements in fuzzy α-clusters.

The squared normalized Euclidean distance (15) and the complement operation (16) were used for construction of the matrix of fuzzy tolerance $T=[\mu_T(x_i,x_j)]$, $i,j=1,\ldots,16$. The technique for detecting the optimal maximal number of elements in fuzzy α-clusters was applied to the matrix by using the measure of separation and compactness of the allotment. The performance of the validity measure is shown in Fig. 3.

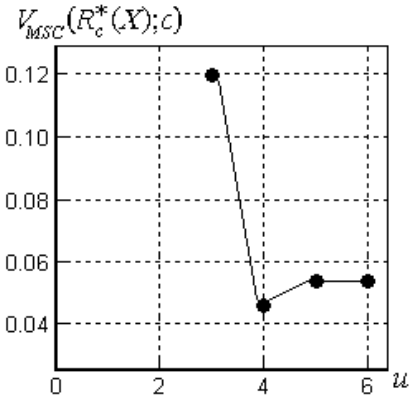

Fig. 3. Plot of the measure of separation and compactness for Sneath and Sokal's data set.

By executing the technique for detecting the optimal maximal number of elements in fuzzy α-clusters, the allotment $R_c^*(X)$ among four fully separated fuzzy clusters, which corresponds to the result, is received for the value $u = 5$ and for the value of tolerance threshold $\alpha = 0.91319$.

Membership functions of four classes are presented in Fig. 4, where membership values of the first class are represented by ○, membership values of the second class are represented by ■, membership values of the third class are represented by ▲, and membership values of the fourth class are represented by □. Values which equal zero are not shown in the Fig. 4.

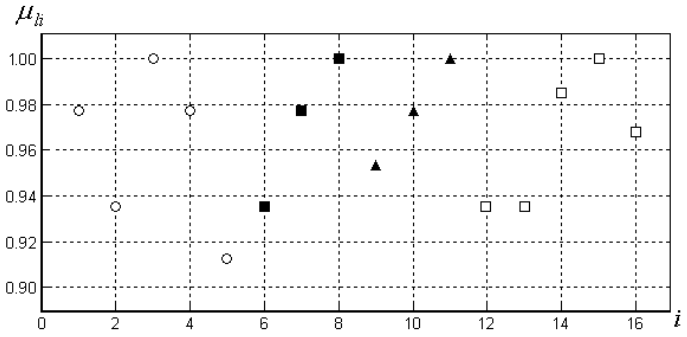

Fig. 4. Membership functions of four fuzzy clusters obtained from the technique for detecting the optimal maximal number of elements in fuzzy α-clusters.

The third object is the typical point of the first fuzzy α-cluster, the eighth object is the typical point of the second fuzzy α-cluster, the eleventh object is the typical point of the third fuzzy α-cluster and the fifteenth object is the typical point of the fourth α-cluster.

The equal result can be obtained by using the linear measure of fuzziness of the allotment and the quadratic measure of fuzziness of the allotment. This fact was shown in [7]. It is should be noted, that the result is equal to the result, obtained from the D-PAFC-algorithm, which was presented in [6]. However, this situation is not met always for other data sets.

V. Concluding Remarks

The procedure for finding the initial minimal number $u_{\min}$ of objects in fuzzy α-clusters is proposed in the paper. The value $u_{\min}$ is the result of application of the procedure to the analyzed data set and the value should be used in the technique for detecting the optimal maximal number of elements in fuzzy α-clusters.

So, the proposed procedure can be considered as the first step of the two-step methodology of constructing the allotment among a priori unknown number of fully separate fuzzy α-clusters with optimal maximal number of elements in the each fuzzy α-cluster. Direct heuristic algorithms of possibilistic clustering for constructing the principal allotment are the essence of the proposed procedure.

The result of application of the two-step methodology to Sneath and Sokal's data set show that the proposed technique is the effective tool for solving the classification problem under a priori uncertainty of the number of fuzzy clusters in the sought allotment, and the number of elements in fuzzy α-clusters.

References

[1] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

[2] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98–110, 1993.

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

[4] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, Wiley, Chichester, 1999.

[5] J. C.Bezdek, J. M. Keller, R. Krishnapuram, and N. R. Pal, *Fuzzy Models and Algorithms for Pattern Recognition and Image Processing*, Springer, New York, 2005.

[6] D. A. Viattchenin, *A Heuristic Approach to Possibilistic Clustering: Algorithms and Applications*, Springer, Heidelberg, 2013.

[7] D. A. Viattchenin, "Heuristic possibilistic clustering for detecting optimal number of elements in fuzzy clusters," *Foundations of Computing and Decision Sciences*, vol. 41, no. 1, pp. 45-76, 2016.

[8] D. A. Viattchenin, A. Yaroma, and A. Damaratski, "A novel direct relational heuristic algorithm of possibilistic clustering," *International Journal of Computer Applications*, vol. 107, no. 18, pp. 15-21, 2014.

[9] D. A. Viattchenin, "A novel heuristic algorithm of possibilistic clustering for given minimal value of the tolerance threshold," *Journal of Information, Control and Management Systems*, vol. 13, no. 2, pp. 161-174, 2015.

[10] D. A. Viattchenin, E. Nikolaenya, and A. Damaratski, "A fuzzy graph-based heuristic algorithm of possibilistic clustering," *Communications on Applied Electronics*, vol. 3, no. 7, pp. 13-23, 2015.

[11] D. A. Viattchenin and A. Damaratski, "Direct heuristic algorithms of possibilistic clustering based on transitive approximation of fuzzy tolerance," *Informatica Economicá*, vol. 17, no.3, pp. 5-15, 2013.

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

[13] M. Walesiak, *Ugólniona Miara Odległości w Statystycznej Analizie Wielowymiarowej*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław, 2002. (in Polish)

[14] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*, Academic Press, New York, 1975.

[15] P. H. A. Sneath and R. Sokal, *Numerical Taxonomy*, Freeman, San Francisco, 1973.