HEURISTIC POSSIBILISTIC CLUSTERING FOR DETECTING OPTIMAL NUMBER OF ELEMENTS IN FUZZY CLUSTERS

Dmitri A. VIATTCHENIN*

Abstract. The paper deals with the problem of discovering fuzzy clusters with optimal number of elements in heuristic possibilistic clustering. The relational clustering procedure using a parameter that controls cluster sizes is considered and a technique for detecting the optimal number of elements in fuzzy clusters is proposed. The effectiveness of the proposed technique is illustrated through numerical examples. Experimental results are discussed and some preliminary conclusions are formulated.

Keywords: possibilistic clustering, fuzzy cluster, allotment, cluster size, cluster validity

1. Introduction

Fuzzy and possibilistic clustering and their necessary formalizations with objective functions are presented in the first subsection. The second subsection includes a brief review of existing algorithms of fuzzy clustering with a variable for controlling cluster size.

1.1. Preliminary remarks

Clustering is the most popular unsupervised machine learning approach in data mining. Clustering is a method which divides an initial set of objects into groups which are referred to as clusters. Elements in same cluster generated by clustering algorithm are considered similar, while elements in different cluster are considered dissimilar. Elements of clusters are objects in data sets. Various methods of cluster analysis were proposed by different researchers. Basic concepts of cluster analysis and a review of the methods are considered by Everitt, Landau, Leese and Stahl [6]. The presented paper develops a heuristic approach to possibilistic clustering which has been proposed in [22]. However, basic notions in fuzzy and possibilistic clustering should be considered in the first place.

---
* Department of Software Information Technology, Belarusian State University of Informatics and Radio-Electronics, P. Brovka St. 6, 220013 Minsk, Belarus, e-mail: viattchenin@mail.ru

There are two kinds of clustering algorithms: hierarchical algorithms versus partitioning algorithms. Hierarchical clustering generates a sequence of nested partitions from the proximity matrix. On the other hand, a single partition of the data is generated by partitioning algorithms.

Clustering algorithms can also be divided into two types: hard versus fuzzy. Hard clustering assigns each input vector to exactly one cluster. On the other hand, in fuzzy clustering, a given element does not necessarily belong to only one cluster but can have varying degrees of memberships to all clusters. Heuristic methods of fuzzy clustering, hierarchical methods of fuzzy clustering and optimization methods of fuzzy clustering were proposed by different researchers.

Fuzzy clustering is more valuable than hard clustering for two reasons. Firstly, the membership degrees give information about the uncertainty of the classification. Secondly, fuzzy clustering can adapt to noisy data and classes that are not well separated.

The most widespread approach in fuzzy clustering is the optimization approach and the traditional optimization methods of fuzzy clustering are based on the concept of fuzzy c-partition. The initial set $X = \{x_1,...,x_n\}$ of n objects represented by the matrix of similarity coefficients, the matrix of dissimilarity coefficients or the matrix of object attributes, should be divided into c fuzzy clusters. Namely, the grade $u_{li} \in [0,1]$, $1 \le l \le c$, $1 \le i \le n$, to which an object x_i belongs to the fuzzy cluster A^l should be determined. For each object x_i, $1 \le i \le n$ the grades of membership should satisfy the conditions of a fuzzy c-partition:

$$\sum_{l=1}^{c} u_{li} = 1, \ 1 \le i \le n, \ 0 \le u_{li} \le 1, \ 1 \le l \le c. \tag{1}$$

In other words, the family of fuzzy sets $P(X) = \{A^l \mid l = \overline{1,c}, c \le n\}$ is the fuzzy c-partition of the initial set of objects $X = \{x_1,...,x_n\}$ if condition (1) is met. Fuzzy c-partition $P(X)$ may be described with the aid of a partition matrix $P_{c \times n} = [u_{li}]$, $l = 1,\ldots,c$, $i = 1,\ldots,n$. The set of all fuzzy c-partitions will be denoted by Π. So, the fuzzy problem formulation in cluster analysis can be defined as the optimization task $Q \to \underset{P(X) \in \Pi}{extr}$ under the constraints (1), where Q is a fuzzy objective function.

Objective function-based fuzzy clustering algorithms can in general be divided into two types: object versus relational. The object data clustering methods can be applied if the objects are represented as points in some multidimensional space. In the relational approach to fuzzy clustering, the problem of the data classification is solved by expressing a relation which quantifies either similarity, or dissimilarity, between pairs of objects.

The best known optimization approach to fuzzy clustering is the method of fuzzy c-means, developed by Bezdek [2]. The FCM-algorithm is based on an iterative optimization of the fuzzy objective function, which takes the form:

$$Q_{FCM}(P,\overline{T}) = \sum_{l=1}^{c}\sum_{i=1}^{n} u_{li}^{\gamma} \left\| x_i - \overline{\tau}^l \right\|^2 , \qquad (2)$$

where u_{li}, $1 \le l \le c$, $1 \le i \le n$ is the membership degree, x_i, $i \in \{1,\ldots,n\}$ is the data point, $\overline{T} = \{\tau^1,\ldots,\tau^c\}$ is the set of fuzzy clusters prototypes and $\gamma > 1$ is the weighting exponent. The FCM-algorithm is very good example of object fuzzy clustering algorithms and the algorithm is the basis of the family of fuzzy clustering algorithms.

Relational clustering algorithms generate a fuzzy c-partition from relational data based on minimization of an objective function. The very popular example of fuzzy relational clustering is the ARCA-algorithm which was proposed by Corsini, Lazzerini, and Marcelloni [4]. The ARCA-algorithm is based on the fuzzy objective function, which takes the form:

$$Q_{ARCA}(P,\overline{T}) = \sum_{l=1}^{c}\sum_{k=1}^{g} u_{lk}^{\gamma} \left(\sqrt{\sum_{j=1}^{n} \left(d(x_i,x_j) - d(x_j,\tau^l) \right)^2} \right)^2 , \qquad (3)$$

where u_{li}, $1 \le l \le c$, $1 \le i \le n$ is the membership degree, x_i, x_j $i,j \in \{1,\ldots,n\}$ are the data points, $\gamma > 1$ is the weighting exponent and $\overline{T} = \{\tau^1,\ldots,\tau^c\}$ is the set of fuzzy clusters prototypes.

If, on the other hand, a condition

$$\sum_{l=1}^{c} u_{li} \ge 1, \ 1 \le i \le n, \ 0 \le u_{li} \le 1, \ 1 \le l \le c , \qquad (4)$$

is met for each object x_i, $1 \le i \le n$, then the corresponding family of fuzzy sets $C(X) = \{A^l \mid l = \overline{1,c}, c \le n\}$ is the fuzzy coverage of the initial set of objects $X = \{x_1,\ldots,x_n\}$. The concept of fuzzy coverage is used mainly in heuristic fuzzy clustering procedures. The FCC-algorithm which was proposed by Chiang, Yue, and Yin [3] is an example of such procedures.

A possibilistic approach to clustering was proposed by Krishnapuram and Keller in [11] and their PCM-algorithm is very effective clustering procedure. The PCM-algorithm is the basis for other algorithms of possibilistic clustering [11]. A concept of possibilistic partition is a basis of possibilistic clustering methods and the membership values μ_{li}, $1 \le l \le c$, $1 \le i \le n$ can be interpreted as the values of typicality degree. For each object x_i, $1 \le i \le n$ the grades of membership should satisfy the conditions of a possibilistic partition:

$$\sum_{l=1}^{c} \mu_{li} > 0, \ 1 \le i \le n, \ 0 \le \mu_{li} \le 1, \ 1 \le l \le c . \qquad (5)$$

So, the family of fuzzy sets $Y(X) = \{A^l \mid l = \overline{1,c}, c \leq n\}$ is the possibilistic partition of the initial set of objects $X = \{x_1, ..., x_n\}$ if condition (5) is met. In particular, the PCM-algorithm is based on an optimization of the objective function

$$Q_{PCM}(Y, \overline{T}) = \sum_{l=1}^{c} \sum_{i=1}^{n} \mu_{li}^{\psi} d^2(x_i, \tau^l) + \sum_{l=1}^{c} \eta_l \sum_{i=1}^{n} (1 - \mu_{li})^{\psi}, \quad (6)$$

under the constraint of possibilistic partition (5) where c is the number of fuzzy clusters A^l, $l = 1, ..., c$ in the possibilistic partition Y, $\mu_{li} \in [0,1]$ is the possibilistic memberships which are typicality degrees, $\overline{\tau}^l \subseteq \mathfrak{R}^m$ is the prototype for fuzzy cluster A^l, the parameter $\psi > 1$ is the analog of the weighting exponent and $d^2(x_i, \tau^l)$ is the squared Euclidean distance:

$$d^2(x_i, \overline{\tau}^l) = \left\| x_i - \overline{\tau}^l \right\|^2. \quad (7)$$

The possibilistic approach to clustering was developed by different researchers. For example, the PCA-algorithm was proposed by Yang and Wu in [27]. Yang and Wu's objective function takes the form:

$$Q_{PCA}(Y, \overline{T}) = \sum_{l=1}^{c} \sum_{i=1}^{n} \mu_{li}^{\psi} d^2(x_i, \tau^l) + \frac{\zeta}{\psi^2 \sqrt{c}} \sum_{l=1}^{c} \sum_{i=1}^{n} (\mu_{li}^{\psi} \log \mu_{li}^{\psi} - \mu_{li}^{\psi}), \quad (8)$$

where $\zeta = \frac{1}{n} \sum_{i=1}^{n} \left\| x_i - \overline{x} \right\|$ with $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $d^2(x_i, \tau^l)$ is the squared Euclidean distance. The parameter ζ is the normalization term that measures the grade of separation of the data set and this parameter can always be fixed as the sample co-variance. The parameter $\sqrt{c}$ is used in (8) to control the steepness degree of the membership functions. The role of the parameter ψ is equal to the fuzzifier in the objective function (6).

On the other hand, a possibilistic clustering method based on a robust approach using Vapnik's [21] ε-intensive estimator, called as the εPCM-algorithm, has been proposed by Łęski [12]. The ε-intensive loss function is:

$$|t|_{\varepsilon} = \begin{cases} 0, & |t| \leq \varepsilon \\ |t| - \varepsilon, & |t| > \varepsilon \end{cases}, \quad (9)$$

where t denotes the value of error and ε denotes the value of a intensivity parameter. Many robust loss functions are described in literature. However, due to its simplicity Vapnik's ε-intensive loss function (9) is of special interest. For example, the well-known absolute error loss function is a special case of (9) for $\varepsilon = 0$. So, if the ε-intensive loss function (9) is put into the possibilistic c-means criterion (6) then the objective function takes the form [12]:

$$Q_{\varepsilon PCM}(Y, \overline{T}) = \sum_{l=1}^{c} \sum_{i=1}^{n} \mu_{li}^{\psi} \left| x_i - \overline{\tau}^l \right|_{\varepsilon} + \sum_{l=1}^{c} \eta_l \sum_{i=1}^{n} (1 - \mu_{li})^{\psi} , \tag{10}$$

where

$$\left| x_i - \overline{\tau}^l \right|_{\varepsilon} = \sum_{t=1}^{m} \left| x_i^t - \overline{\tau}^{(t)l} \right|_{\varepsilon} , \tag{11}$$

Therefore, the purpose of the classification task is to obtain the solutions $Y(X)$ and $\tau^1, \ldots, \tau^c$ which minimize (10) under the constraints (5).

Some other possibilistic clustering procedures have also been developed. In particular, the relational RPCM-algorithm of possibilistic clustering has been proposed by De Cáceres, Oliva and Font [5], the PGCM-algorithm has been introduced by Ménard, Courboulay and Dardignac [14], and the EPCM-algorithm has been proposed by Xie, Wang and Chung [26]. Many fuzzy and possibilistic clustering algorithms could be found in the corresponding books [2], [8], [15], [18].

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 conform to a given definition. Mandel [13] note that such algorithms are called algorithms of direct classification or direct clustering algorithms. Direct heuristic algorithms of fuzzy clustering are simple and very effective in many cases. So, a heuristic approach to possibilistic clustering has been also proposed in [22].

1.2. Objective functions with variables for controlling cluster size

There is a problem to separate a dense cluster and a sparse cluster for which density or cluster size has to be considered. For the purpose, some objective functions and corresponding algorithms are considered in [15]. Let us remind these objective functions with variables for controlling cluster size.

Using an additional variable that controls cluster sizes, or, in other words, cluster volumes, is a natural idea. In the first place, the objective function

$$Q_{FCMA}(P, \overline{T}, R) = \sum_{l=1}^{c} \sum_{i=1}^{n} R_l^{1-\gamma} u_{li}^{\gamma} d^2(x_i, \overline{\tau}^l) , \tag{12}$$

is considered in [15], where u_{li}, $l = 1, \ldots, c$, $i = 1, \ldots, n$ is the membership degree, x_i, $i \in \{1, \ldots, n\}$ is the data point, $\overline{T} = \{\overline{\tau}^1, \ldots, \overline{\tau}^c\}$ is the set fuzzy clusters prototypes, $\gamma > 1$ is the weighting exponent, and $d^2(x_i, \overline{\tau}^l)$ is the squared Euclidean distance between x_i and $\overline{\tau}^l$ (7). The variable $R = (R_1, \ldots, R_c)$ controls cluster sizes. The constraint for R is

$$\left\{ R = (R_1, \ldots, R_c) : \sum_{l=1}^{c} R_l = 1; R_l \geq 0, 1 \leq l \leq c \right\}. \tag{13}$$

In the second place, the objective function

$$Q_{FCMAS}(P, \overline{T}, R, C) = \sum_{l=1}^{c} \sum_{i=1}^{n} R_l^{1-\gamma} u_{li}^{\gamma} d^2(x_i, \overline{\tau}^l), \tag{14}$$

is considered in [15], where $d^2(x_i, \overline{\tau}^l)$ is the squared Mahalanobis distance

$$d^2(x_i, \overline{\tau}^l) = (x_i - \overline{\tau}^l)^{\mathrm{T}} C_l^{-1} (x_i - \overline{\tau}^l), \tag{15}$$

and C_l is a positive definite matrix with the constraint $\det(C_l) = \vartheta_l$, where $\vartheta_l > 0$ is a fixed parameter.

In the third place, the method based on the Kullback-Leibler information is considered in [15]. The method uses the next objective function:

$$Q_{KL}(P, \overline{T}, R, C) = \sum_{l=1}^{c} \sum_{i=1}^{n} u_{li} d^2(x_i, \overline{\tau}^l) + \sum_{l=1}^{c} \sum_{i=1}^{n} u_{li} \left(\lambda \log \frac{u_{li}}{R_l} + \log(\det(C_l))^{1/2} \right), \tag{16}$$

where $\lambda > 0$ and $d^2(x_i, \overline{\tau}^l)$ is the squared Mahalanobis distance (15).

Different aspects of controlling cluster sizes are discussed in the literature. In particular, the discussion of objective functions (12), (14), and (16) could be found in [7], [10], [16].

The most important problem of fuzzy clustering is neither the choice of the numerical procedure nor the distance to use but concerns the number c of fuzzy clusters to look for. Really, lacking in a priori knowledge of the data structure, there is no reason to choose a particular value of c and one must find a way to measure the acceptance with which cluster structure has been identified by a clustering procedure. This is the so-called cluster validity problem.

The classical approach to cluster validity for fuzzy clustering is based on directly evaluating the fuzzy c-partition. Measures of cluster validity can be used for the purpose. Many authors have proposed several measures of cluster validity associated with fuzzy c-partitions. These validity measures are considered in [2], [8], [15]. However, Pedrycz [17] note that the behavior of cluster validity indexes has not been theoretically justified, but simulation experiments confirmed their utility. It is should be note, that the relationship between the cluster validity problem and the problem of controlling cluster sizes is not considered in [15].

The main goal of this paper is a consideration of the relationship between the optimal maximal number of elements in each fuzzy cluster, and the number of fuzzy clusters in the framework of the heuristic approach to possibilistic clustering. For the purpose, 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. The technique is based on the

direct relational D-AFC(u)-algorithm of possibilistic clustering and corresponding cluster validity measures.

So, the content of this paper is as follows: in the second section basic definitions of the heuristic approach to possibilistic clustering are outlined, methods of the data preprocessing are considered, and validity measures for relational heuristic algorithms of possibilistic clustering are presented, in the third section a general plan of the relational heuristic D-AFC(u)-algorithm of possibilistic clustering is given and a technique for detecting the optimal maximal number of elements in fuzzy clusters is provided, in the fourth section illustrative examples of the proposed technique application to well-known data sets are presented, in the fifth section results of numerical experiments are discussed and some preliminary conclusions are formulated.

2. A heuristic approach to possibilistic clustering

Basic concepts of a heuristic approach to possibilistic clustering are considered in the first subsection. A brief review of heuristic possibilistic clustering procedures is presented in the second subsection of the section. The third subsection includes a brief review of methods of the data pre-processing. Methods for evaluating clustering results and validity measures for the basic heuristic algorithm of possibilistic clustering are described in the fourth subsection.

2.1. Basic definitions

Let us remind basic concepts of a heuristic method of possibilistic clustering which was proposed in [22]. The essence of the heuristic approach to possibilistic clustering is that the sought clustering structure of the set of objects is formed based directly on the formal definition of fuzzy cluster and possibilistic memberships are determined also directly from the values of the pair wise similarity of objects.

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 $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha\}$, $l \in [1,n]$ is fuzzy α-cluster or, simply, fuzzy cluster. So $A^l_{(\alpha)} \subseteq A^l$, $\alpha \in (0,1]$, $A^l \in \{A^1,...,A^n\}$ and μ_{li} is the membership degree of the element $x_i \in X$ for some fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$. 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

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A_\alpha^l \\ 0, & otherwise \end{cases} \qquad (17)$$

where an α-level $A_\alpha^l = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$ of a fuzzy set A^l is the support of the fuzzy cluster $A_{(\alpha)}^l$. So, condition $A_\alpha^l = Supp(A_{(\alpha)}^l)$ is met for each fuzzy cluster $A_{(\alpha)}^l$, $\alpha \in (0,1]$, $l \in [1, n]$. Membership degree can be interpreted as a degree of typicality of n element to a fuzzy cluster.

Let T is a fuzzy tolerance on X, where X is the set of objects, and $\{A_{(\alpha)}^1, ..., A_{(\alpha)}^n\}$ is the family of fuzzy clusters for some $\alpha \in (0,1]$. The point $\tau_e^l \in A_\alpha^l$, for which

$$\tau_e^l = \arg\max_{x_i} \mu_{li}, \ \forall x_i \in A_\alpha^l, \qquad (18)$$

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

Let $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1, c}, 2 \leq c \leq n, \alpha \in (0,1]\}$ be a family of fuzzy 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 $X = \{x_1, ..., x_n\}$. If a condition

$$\sum_{l=1}^c \mu_{li} > 0, \ \forall x_i \in X \qquad (19)$$

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

The condition (19) requires that every object x_i, $i = 1, ..., n$ must be assigned to at least one fuzzy cluster $A_{(\alpha)}^l$, $l = \overline{1, c}$, $c \leq n$ with the membership degree higher than zero. The condition $2 \leq c \leq n$ requires that the number of fuzzy clusters in each allotment $R_{c(z)}^\alpha(X)$ must be equal or 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 (19) is similar to the definition of the possibilistic partition (5). So, the allotment among fuzzy clusters can be considered as

the possibilistic partition, and fuzzy clusters in the sense of definition (17) are elements of the possibilistic partition. However, the concept of allotment will be used in further considerations.

Allotment $R_I^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,n},\, \alpha \in (0,1]\}$ of the set of objects among n fuzzy clusters for some tolerance threshold $\alpha \in (0,1]$ is the initial allotment of the set $X = \{x_1,...,x_n\}$. 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 α-level fuzzy sets $A_{(\alpha)}^l$, $l = \overline{1,c}$, $\alpha \in (0,1]$ are fuzzy clusters. These fuzzy clusters constitute an initial allotment for some tolerance threshold α and they can be considered as clustering components.

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

$$\bigcup_{l=1}^{c} A_\alpha^l = X \tag{20}$$

and a condition

$$card(A_\alpha^l \cap A_\alpha^m) = 0,\ \forall A_{(\alpha)}^l, A_{(\alpha)}^m,\ l \ne m,\ \alpha \in (0,1] \tag{21}$$

are met for all fuzzy clusters $A_{(\alpha)}^l$, $l = \overline{1,c}$ of some allotment $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c},\, c \le n\}$ for a value $\alpha \in (0,1]$, then the allotment is the allotment among fully separate fuzzy clusters.

Fuzzy clusters in the sense of definition (17) can have an intersection area. If the intersection area of any pair of different fuzzy clusters is an empty set, then conditions (20) and (21) are met and fuzzy clusters are called fully separate fuzzy clusters. Otherwise, fuzzy clusters are called particularly separate fuzzy clusters and $w \in \{0,...,n\}$ is the maximum number of elements in the intersection area of different fuzzy clusters. For $w = 0$ fuzzy clusters are fully separate fuzzy clusters. Thus, the conditions (20) and (21) can be generalized for a case of particularly separate fuzzy clusters. So, a condition

$$\sum_{l=1}^{c} card(A_\alpha^l) \ge card(X), \quad \forall A_{(\alpha)}^l \in R_{c(z)}^\alpha(X),$$
$$\alpha \in (0,1], \quad card(R_{c(z)}^\alpha(X)) = c \tag{22}$$

and a condition

$$card(A_\alpha^l \cap A_\alpha^m) \le w,\ \forall A_{(\alpha)}^l, A_{(\alpha)}^m,\ l \ne m,\ \alpha \in (0,1], \tag{23}$$

are generalizations of conditions (20) and (21). Obviously, if $w = 0$ in conditions (22) and (23) then conditions (20) and (21) are met. The adequate allotment $R_{c(z)}^{\alpha}(X)$ for some value of tolerance threshold $\alpha \in (0,1]$ is a family of fuzzy clusters which are elements of the initial allotment $R_{I}^{\alpha}(X)$ for the value of α and the family of fuzzy clusters should satisfy the conditions (22) and (23). So, the construction of adequate allotments $R_{c(z)}^{\alpha}(X) = \{A_{(\alpha)}^{l} \mid l = \overline{1,c}, c \le n\}$ for every α is a trivial problem of combinatorics.

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

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, $B = \{R_{c(z)}^{\alpha}(X)\}$, 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 $B = \{R_{c(z)}^{\alpha}(X)\}$ of adequate allotments must be made on the basis of evaluation of allotments. In particular, the criterion

$$F(R_{c(z)}^{\alpha}(X),\alpha) = \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c, \qquad (24)$$

where c is the number of fuzzy clusters in the allotment $R_{c(z)}^{\alpha}(X)$ and $n_l = card(A_{\alpha}^{l})$, $A_{(\alpha)}^{l} \in R_{c(z)}^{\alpha}(X)$ 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 (24) corresponds to the best allotment of objects among c fuzzy clusters. So, the classification problem can be characterized formally as determination of the solution $R_{c}^{*}(X)$ satisfying

$$R_{c}^{*}(X) = \arg \max_{R_{c(z)}^{\alpha}(X) \in B} F(R_{c(z)}^{\alpha}(X),\alpha). \qquad (25)$$

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 and detection of fixed or unknown number c of fuzzy clusters can be considered as the aim of classification.

2.2. 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 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 for the prototype-based algorithms.

In particular, the group of direct relational heuristic algorithms of possibilistic clustering includes

- D-AFC(c)-algorithm: using the construction of the allotment among given number c of partially separate fuzzy clusters;
- D-PAFC-algorithm: using the construction of the principal allotment among an unknown minimal number of at least c fully separate fuzzy clusters;
- D-AFC-PS(c)-algorithm: using the partially supervised construction of the allotment among given number c of partially separate fuzzy clusters.

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

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

It should be noted that these direct prototype-based heuristic possibilistic clustering algorithms are based on a transitive closure of an initial fuzzy tolerance relation. 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 [23]. 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.

2.3. A note 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

$$\hat{\rho}_{n\times n} = \begin{pmatrix} \hat{\rho}_{11} & \hat{\rho}_{12} & \cdots & \hat{\rho}_{1n} \\ \hat{\rho}_{21} & \hat{\rho}_{22} & \cdots & \hat{\rho}_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \hat{\rho}_{n1} & \hat{\rho}_{n2} & \cdots & \hat{\rho}_{nn} \end{pmatrix}, \tag{26}$$

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.

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 $\hat{X}_{n \times m_1} = [\hat{x}_i^{t_1}]$, $i = 1, \ldots, n$, $t_1 = 1, \ldots, m_1$ and the data are called sometimes the two-way data [18]. Let $X = \{x_1, \ldots, x_n\}$ is the set of objects. So, the two-way data matrix can be represented as follows:

$$\hat{X}_{n \times m_1} = \begin{pmatrix} \hat{x}_1^1 & \hat{x}_1^2 & \ldots & \hat{x}_1^{m_1} \\ \hat{x}_2^1 & \hat{x}_2^2 & \ldots & \hat{x}_2^{m_1} \\ \ldots & \ldots & \ldots & \ldots \\ \hat{x}_n^1 & \hat{x}_n^2 & \ldots & \hat{x}_n^{m_1} \end{pmatrix}, \quad (27)$$

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

The matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, n$ is the matrix of initial data for the relational heuristic algorithms of possibilistic clustering. However, the data can be presented as a matrix of attributes $\hat{X}_{n \times m_1} = [\hat{x}_i^{t_1}]$, $i = 1, \ldots, n$, $t_1 = 1, \ldots, m_1$, 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 (27), by choosing a suitable metric to measure similarity.

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

$$x_i^{t_1} = \frac{\hat{x}_i^{t_1}}{\max\limits_{i} \hat{x}_i^{t_1}}. \quad (28)$$

The data normalization method (28) 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

$$x_i^{t_1} = \frac{\hat{x}_i^{t_1} - \min\limits_{i} \hat{x}_i^{t_1}}{\max\limits_{i} \hat{x}_i^{t_1} - \min\limits_{i} \hat{x}_i^{t_1}}. \quad (29)$$

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. Of course, 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 [25].

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 $X = \{x_1, \ldots, x_n\}$ are:

- The normalized Hamming distance:

$$l(x_i, x_j) = \frac{1}{m_1} \sum_{t_1=1}^{m_1} |\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1})|, \; i, j = 1, \ldots, n, \tag{30}$$

- The normalized Euclidean distance:

$$e(x_i, x_j) = \sqrt{\frac{1}{m_1} \sum_{t_1=1}^{m_1} \left(\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1})\right)^2}, \; i, j = 1, \ldots, n, \tag{31}$$

- The squared normalized Euclidean distance:

$$\varepsilon(x_i, x_j) = \frac{1}{m_1} \sum_{t_1=1}^{m_1} \left(\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1})\right)^2, \; i, j = 1, \ldots, n. \tag{32}$$

These distances are considered by Kaufmann [9] in detail. 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

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j), i, j = 1, \ldots, n, \tag{33}$$

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

2.4. Evaluating the fuzzy clusters and validity measures

The result of classification must be interpreted from essential positions. Some formal criteria can be useful for the aim. For example, most appropriate distance between fuzzy sets for the data preprocessing can be selected on a basis of the evaluation of the results of classification. A problem of the evaluation of fuzzy clusters was considered in [22].

The qualitative inspection of fuzzy clustering results can be done, e.g., with a linear index of fuzziness or a quadratic index of fuzziness, used for evaluation of fuzziness degree of fuzzy clusters. These two indexes are considered by Kaufmann [9]. So, a modification of the linear index of fuzziness is defined in [22] as

$$I_L(A_{(\alpha)}^l) = \frac{2}{n_l} \cdot d_H(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l), \tag{34}$$

where $n_l = card(A_\alpha^l)$, $A_{(\alpha)}^l \in R_c^*(X)$ is the number of objects in the fuzzy cluster $A_{(\alpha)}^l$ and $d_H(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l)$ is the Hamming distance

$$d_H(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l) = \sum_{x_i \in A_\alpha^l} \left| \mu_{li} - \mu_{\underline{A}_{(\alpha)}^l}(x_i) \right|, \tag{35}$$

between the fuzzy cluster $A_{(\alpha)}^l$ and the crisp set $\underline{A}_{(\alpha)}^l$ nearest to the fuzzy cluster $A_{(\alpha)}^l$. The membership function of the crisp set $\underline{A}_{(\alpha)}^l$ can be defined as

$$\mu_{\underline{A}_{(\alpha)}^l}(x_i) = \begin{cases} 0, & \mu_{A_{(\alpha)}^l}(x_i) \le 0.5 \\ 1, & \mu_{A_{(\alpha)}^l}(x_i) > 0.5 \end{cases}, \ \forall x_i \in A_\alpha^l, \tag{36}$$

where $\alpha \in (0,1]$.

The modified quadratic index of fuzziness is defined in [22] as

$$I_Q(A_{(\alpha)}^l) = \frac{2}{\sqrt{n_l}} \cdot d_E(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l), \tag{37}$$

where $n_l = card(A_\alpha^l)$, $A_{(\alpha)}^l \in R_c^*(X)$ and $d_E(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l)$ is the Euclidean distance

$$d_E(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l) = \sqrt{\sum_{x_i \in A_\alpha^l} \left(\mu_{li} - \mu_{\underline{A}_{(\alpha)}^l}(x_i) \right)^2}, \tag{38}$$

between the fuzzy cluster $A_{(\alpha)}^l$ and the crisp set $\underline{A}_{(\alpha)}^l$ which is defined by formula (36).

Indexes (34) and (37) show the degree of fuzziness of fuzzy clusters which are elements of the allotment $R_c^*(X)$. Obviously, that the condition $I_L(A_{(\alpha)}^l) = I_Q(A_{(\alpha)}^l) = 0$ is met for a crisp set $A_{(\alpha)}^l \in R_c^*(X)$. Otherwise, if $\mu_{li} = 0.5$, $\forall x_i \in A_\alpha^l$ then fuzzy cluster $A_{(\alpha)}^l \in R_c^*(X)$ is a maximally fuzzy set and the condition $I_L(A_{(\alpha)}^l) = I_Q(A_{(\alpha)}^l) = 1$ is met.

The density of fuzzy cluster was defined in [22] as follows:

$$D(A_{(\alpha)}^l) = \frac{1}{n_l} \sum_{x_i \in A_\alpha^l} \mu_{li}, \tag{39}$$

where $n_l = card(A_\alpha^l)$, $A_{(\alpha)}^l \in R_c^*(X)$ and membership degree μ_{li} is defined by formula (17). It is obvious that condition

$$0 < D(A_{(\alpha)}^l) \le 1, \tag{40}$$

is met for each fuzzy cluster $A_{(\alpha)}^l$ in $R_c^*(X)$. Moreover, $D(A_{(\alpha)}^l) = 1$ for a crisp set $A_{(\alpha)}^l \in R_c^*(X)$ for any tolerance threshold α, $\alpha \in (0,1]$. The density of fuzzy cluster shows an average membership degree of elements of a fuzzy cluster.

Let us remind the cluster validity problem for relational heuristic algorithms of possibilistic clustering. The number c of fuzzy clusters and their compactness are contradictory purposes of the classification of n objects. If compact classes are searched, the most appropriate solution can be obtained with n classes of one object. Obviously, that the solution is not useful. So, the number c of fuzzy clusters must be determined under consideration of the conditions: firstly, the number of fuzzy clusters c in the sought allotment $R_c^*(X)$ must be as possible as less, and, secondly, the membership function of fuzzy clusters of some allotment among c fuzzy clusters must be sharper than the membership function of fuzzy clusters of allotments for other numbers of fuzzy clusters.

Let $R_c^*(X)$ be the allotment which is corresponds to the result of classification for the given number c of fuzzy clusters and the and R^c be the set of all allotments $R_c^*(X)$ among c, $c \in \{2,\dots,n\}$ fuzzy clusters. A cluster validity measure can be defined as a mapping $V : R^c \mapsto \Re$ which can be used to rank the validity of various allotments $R_c^*(X)$. Validity measures can be obtained from the indexes which are defined in previous considerations.

The fuzziness of the allotment $R_c^*(X)$ among c fuzzy clusters can be evaluated as the sum of indexes of fuzziness of fuzzy clusters of the allotment $R_c^*(X)$. So, the linear measure of fuzziness of the allotment must be based on the formula (30), and the measure can be defined as follows:

$$V_{LMF}\left(R_c^*(X);c\right) = \sum_{A_{(\alpha)}^l \in R_c^*(X)} \left(I_L(A_{(\alpha)}^l)\right) = \sum_{A_{(\alpha)}^l \in R_c^*(X)} \left(\frac{2}{n_l} \cdot d_H(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l)\right), \tag{41}$$

where $I_L(A_{(\alpha)}^l)$ is the modified linear index of fuzziness (34).

On the other hand, the quadratic measure of fuzziness of the allotment can be defined on the analogy of the linear measure of fuzziness (41):

$$V_{QMF}\left(R_c^*(X);c\right) = \sum_{A_{(\alpha)}^l \in R_c^*(X)} \left(I_Q(A_{(\alpha)}^l)\right) = \sum_{A_{(\alpha)}^l \in R_c^*(X)} \left(\frac{2}{\sqrt{n_l}} \cdot d_E(A_{(\alpha)}^l, \underline{A}_{(\alpha)}^l)\right), \tag{42}$$

where $I_Q(A^l_{(\alpha)})$ is the modified quadratic index of fuzziness (37).

Using $V_{LMF}(R^*_c(X);c)$ or $V_{QMF}(R^*_c(X);c)$, the optimal number c of fuzzy clusters can be obtained by maximizing the index value.

The density of fuzzy cluster (39) can be considered as the basis for a validity measure [22]. So, the measure of separation and compactness of the allotment can be defined in the following way:

$$V_{MSC}(R^*_c(X);c) = \frac{\sum\limits_{A^l_{(\alpha)} \in R^*_c(X)} D(A^l_{(\alpha)})}{c} + \frac{c}{n}\sum\limits_{x_j \in \Theta} \mu_{lj} - \alpha , \qquad (43)$$

where Θ is a set of elements x_j, $j \in \{1,\ldots,n\}$ in all intersection areas of different fuzzy clusters.

The measure of separation and compactness of the allotment $V_{MSC}(R^*_c(X);c)$ increases when c is closer to n. Thus, optimum value of c is obtaining by minimizing $V_{MSC}(R^*_c(X);c)$ over $c = c_{\min},\ldots,c_{\max}$ where $2 \le c_{\min}$ and $c_{\max} < n$.

3. A technique for detecting the optimal number of elements in fuzzy clusters

The relational heuristic D-AFC(u)-algorithm of possibilistic clustering is described in the first subsection. The second subsection includes the detail consideration of a technique for detecting the optimal maximal number of elements in fuzzy clusters.

3.1. The D-AFC(u)-algorithm

Let us remind the D-AFC(u)-algorithm which was proposed in [24]. A new parameter for direct relational heuristic algorithms of possibilistic clustering is a basis for the relational clustering procedure.

An analyst can determine the maximal number u of elements in a fuzzy cluster. If $1 \le u < n$ is a maximal number of elements in a fuzzy cluster, then $1 \le n_l \le u$, $\forall l = \overline{1,c}$, where $n_l = card(A^l_\alpha)$, $A^l_\alpha = Supp(A^l_{(\alpha)})$ for each fuzzy cluster $A^l_{(\alpha)}$, $l = \overline{1,c}$, $\alpha \in (0,1]$. So, parameter u can be considered as the parameter that controls cluster sizes.

Thus, the classification problem can be formulated as follows: detection of an unknown number c of partially separated fuzzy clusters with given maximal number of elements $1 \le u < n$ in every class can be considered as the aim of classification. So, the corresponding D-AFC(u)-algorithm for detecting the allotment among fuzzy clusters with

given maximal number of elements u in every class is an eleven-step procedure of classification.

1.   Calculate α-level values of the fuzzy tolerance T and construct the ordered sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \leq 1$ of α-levels; set $\ell := 0$;

2.   Construct the initial allotment $R_I^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1, n}\}$, $\alpha = \alpha_\ell$;

3.   The following condition is checked:

> **if** for some fuzzy cluster $A_{(\alpha)}^l \in R_I^\alpha(X)$, $l \in \{1, \ldots, n\}$, $\alpha = \alpha_\ell$ the condition $n_l = n$ is met
>
> **then** set $\ell := \ell + 1$ and go to step 2;

4.   Construct the set U of possible clustering components as follows:

> **if** for some fuzzy cluster $A_{(\alpha)}^l \in R_I^\alpha(X)$, $l \in \{1, \ldots, n\}$, $\alpha = \alpha_\ell$ the condition $n_l \leq u$ is met
>
> **then** $A_{(\alpha)}^l \in U$
>
> **else** $A_{(\alpha)}^l \notin U$;

5.   The following condition is checked:

> **if** condition $card(U) \geq 2$ **and** condition $\sum_{l=1}^{c} card(A_\alpha^l) \geq card(X)$,
>
> $\forall A_{(\alpha)}^l \in U$, $\alpha = \alpha_\ell$ are not met
>
> **then** set $\ell := \ell + 1$ and go to step 2
>
> **else** go to step 6;

6.   Set $w := 0$;

7.   Check if it is possible to construct allotments $R_{c(z)}^\alpha(X) = \{A_{(\alpha)}^l \mid A_{(\alpha)}^l \in U\}$, $c \leq n$, which satisfy conditions (18) and (19) for the value $\alpha = \alpha_\ell$;

8.   The following condition is checked:

> **if** allotments $R_{c(z)}^\alpha(X)$ satisfying conditions (22) and (23) are not constructed
>
> **then** set $w := w + 1$ and go to step 7
>
> **else** go to step 9;

9.   Construct the class of possible solutions of the classification problem $B(u) = \left\{ R_{c(z)}^\alpha(X) \right\}$, $\alpha = \alpha_\ell$ for all allotments $R_{c(z)}^\alpha(X)$ which were obtained on the step 7;

10.   Calculate the value of the criterion (24) for every allotment $R_{c(z)}^\alpha(X) \in B(u)$;

11.   The result $R_c^*(X)$ of classification is formed as follows:

if for some unique allotment $R^{\alpha}_{c(z)}(X) \in B(u)$ the condition (25) is met

then the allotment is the result of classification $R^{*}_{c}(X)$

else   **if** the condition $u > 1$ is met

           then set $u := u - 1$ and go to step 7.

The allotment $R^{*}_{c}(X)$ among unknown number c partially separate fuzzy clusters with determined sizes and the corresponding value of tolerance threshold $\alpha \in (0,1]$ are results of classification obtained from the D-AFC(u)-algorithm.

3.2. The proposed technique

The D-AFC(u)-algorithm can be considered as an appropriate tool for detecting the optimal number of elements in fuzzy clusters of the constructed allotment. For the purpose, validity measures should be used because the number of fuzzy clusters in the sought allotment depends on the number of elements in each fuzzy cluster. In other words, if $n_l \to 1$, $l = \overline{1, c}$ then $c \to n$. Let $V(R^{*(k)}_{c}(X); c)$ be a general notation for validity measures (41) – (43). The value u can vary in the interval $[u_{\min}, u_{\max}]$.

Optimal number of elements in fuzzy clusters depends on next criteria:

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

So, the proposed technique for detecting the optimal maximal number of elements in fuzzy 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 $R^{*(k)}_{c}(X)$ will be constructed;
3.   Calculate the value of some validity measure $V(R^{*(k)}_{c}(X); c)$
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 $\left| V(R^{*(k+1)}_{c}(X); c) - V(R^{*(k)}_{c}(X); c) \right| = 0$ is met

then the value u_k is the optimal number of elements in each fuzzy cluster $A^l_{(\alpha)}$, $l = 1, \ldots, c$ of the constructed allotment $R^{*(k)}_c(X)$ and stop

**else** set $k := k + 1$, $u_{k+1} := u_k + 1$ and go to step 2;

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

4. Illustrative examples

The first subsection includes the Tamura's relational data description and results of their processing by the D-AFC(u)-algorithm using the proposed technique. In the second subsection the Sneath and Sokal's two-dimensional data set is used for the testing of the proposed technique. An application of the proposed technique to the classification problem for the Anderson's Iris data set is considered in the third subsection of the section.

4.1. The Tamura's portrait data set

Let us consider an application of proposed technique 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 in [20]. The number of portraits was equal to 16 and the real portrait assignment among three classes is presented in Figure 1.
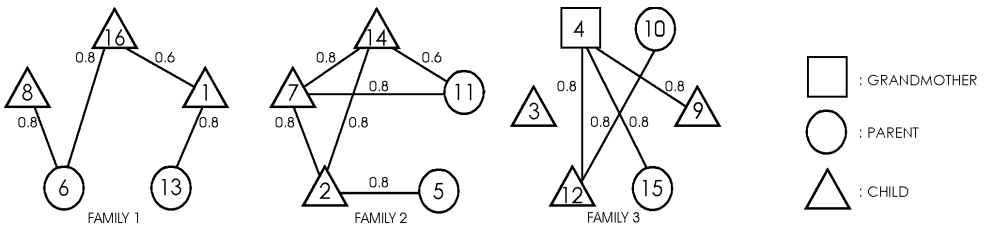
**Figure 1.   Real portraits classification**

The data were originally analyzed in order to identify families with the technique of first transforming the matrix of a fuzzy tolerance into a matrix of a fuzzy similarity relation and then taking an appropriate α-level of the fuzzy similarity relation [20]. The best partition proved to be obtained with α-level equal to 0.6. The partition identified the following three families $A^1 = \{x_1, x_6, x_8, x_{13}, x_{16}\}$, $A^2 = \{x_2, x_5, x_7, x_{11}, x_{14}\}$ and $A^3 = \{x_4, x_9, x_{10}, x_{12}, x_{15}\}$. However, person x_3 is not a member of any of the three families.

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

Table 1.   **The matrix of subjective similarities**

| 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

The proposed technique for detecting the optimal number of elements in fuzzy clusters was applied to the matrix of fuzzy tolerance for $u \in [u_{\min} = 3, u_{\max} = 8]$ using validity measures (41) – (43). The performance of validity measures is shown in Figures 2 – 4.
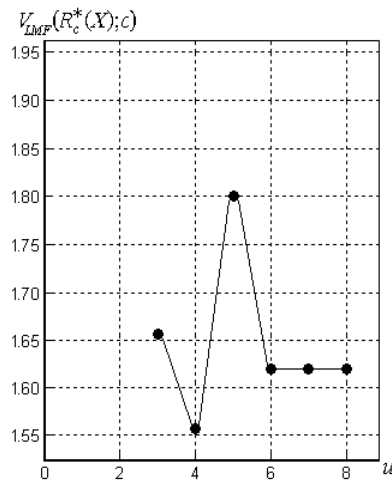
**Figure 2.**   **Plot of the linear measure of fuzziness for Tamura's data set**
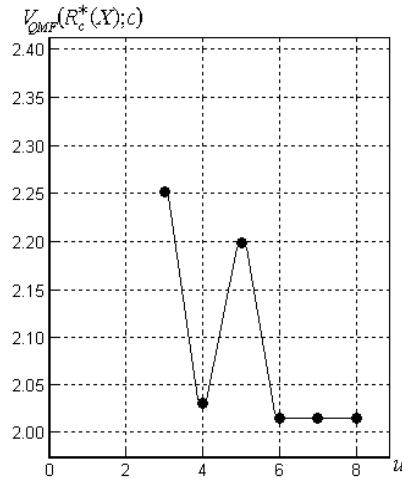
Figure 3.   Plot of the quadratic measure of fuzziness for Tamura's data set
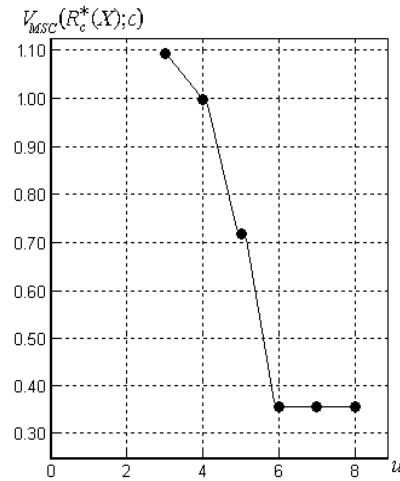
Figure 4.   Plot of the measure of separation and compactness for Tamura's data set

So, the optimal number of elements in each fuzzy cluster is equal 6. The result corresponds to the allotment $R_c^*(X)$ among four fully separated fuzzy clusters. Membership functions of four fuzzy clusters obtained by using the D-AFC(u)-algorithm for $u = 6$ are shown in Figure 5, 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 ∇.
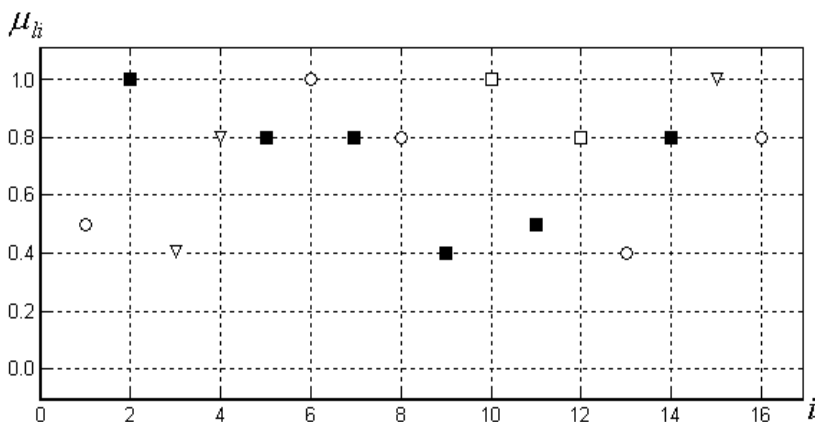
Figure 5.   Membership functions of four fully separated fuzzy clusters

By executing the D-AFC(u)-algorithm for $u = 6$, the following result is obtained: the first class is composed of 5 elements, all belonging to Family 1; the second class contains 2 elements, all from Family 3; the third class is formed by 6 elements, where five elements correspond to Family 2 and one element corresponds to Family 3, and the fourth class consists of 3 elements, all belonging to 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 fuzzy cluster which corresponds to the first class is maximal for the sixth object and is equal one. That is why the sixth object is the typical point of the first fuzzy cluster. The membership value of the tenth object is equal one for the second fuzzy cluster and the tenth object is the typical point of the second fuzzy cluster. The membership value of the second object is equal one for the fuzzy cluster which corresponds to the third class. So, the second object is the typical point of the third fuzzy cluster. The membership value of the fifteenth object is equal one for the fourth fuzzy cluster. Thus, the fifteenth object is the typical point of the fuzzy cluster which corresponds to the fourth class.

So, the results of the proposed technique for detecting the optimal maximal number of elements in fuzzy clusters seem to be appropriate.

4.2.   The Sneath and Sokal's two-dimensional data set

An application of the proposed technique to the data processing can be explained also by the second example. For the purpose, the Sneath and Sokal's two-dimensional data set [19] was selected. The artificial data set is shown in Figure 6.
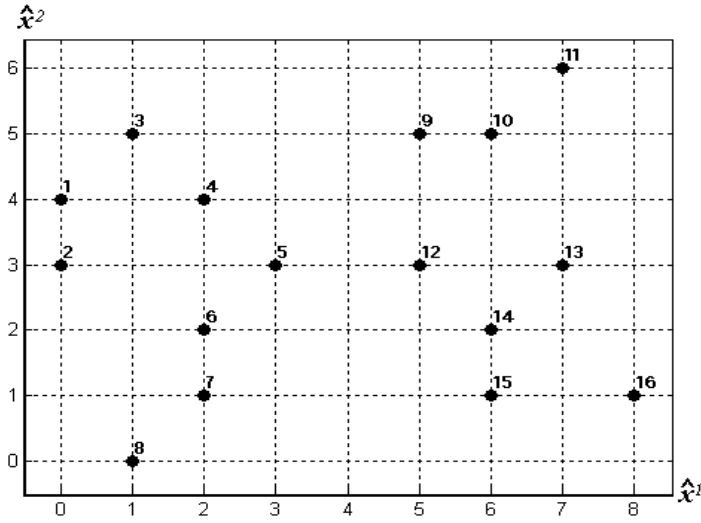
Figure 6.   The Sneath and Sokal's data set

The initial data set was preprocessed by using formulae (28), (32), (33). So, the matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, 16$ was obtained. The proposed technique for detecting the optimal number of elements in fuzzy clusters was also applied to the matrix of fuzzy tolerance for $u \in [u_{\min} = 3, u_{\max} = 8]$ using validity measures (41) – (43). The performance of validity measures is shown in Figures 7 – 9.
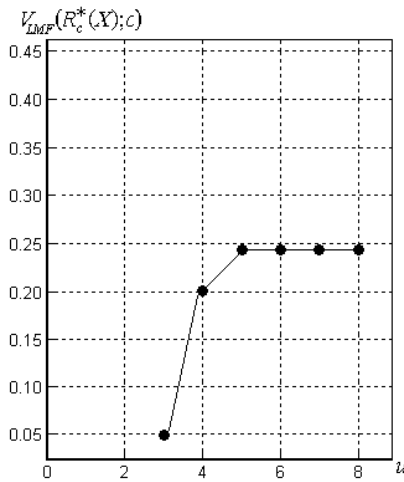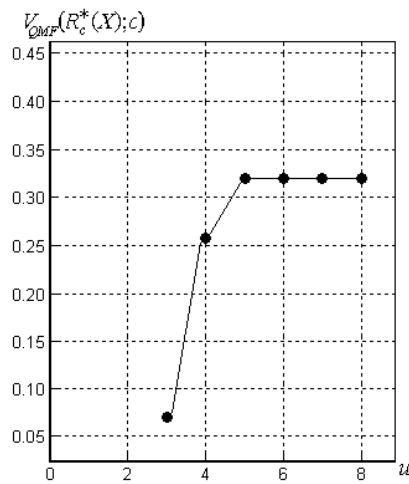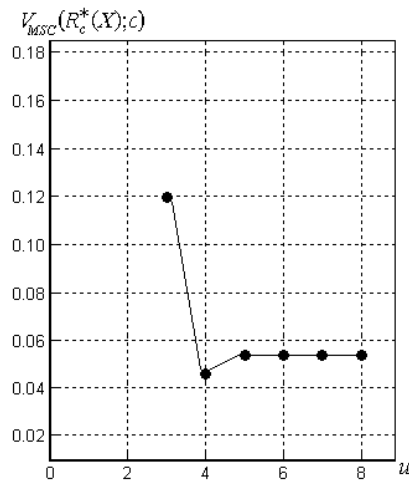

Figure 7.   Plot of the linear measure of fuzziness for Sneath and Sokal's data set

$$V_{QMF}(R_c^*(X);c)$$

Figure 8. **Plot of the quadratic measure of fuzziness for Sneath and Sokal's data set**

$$V_{MSC}(R_c^*(X);c)$$

Figure 9. **Plot of the measure of separation and compactness for Sneath and Sokal's data set**

By executing the proposed technique, the allotment $R_c^*(X)$ among four fully separated fuzzy clusters, which corresponds to the result, is received for the value $u = 5$.

Membership functions of four classes are presented in Figure 10, 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 Figure 10.
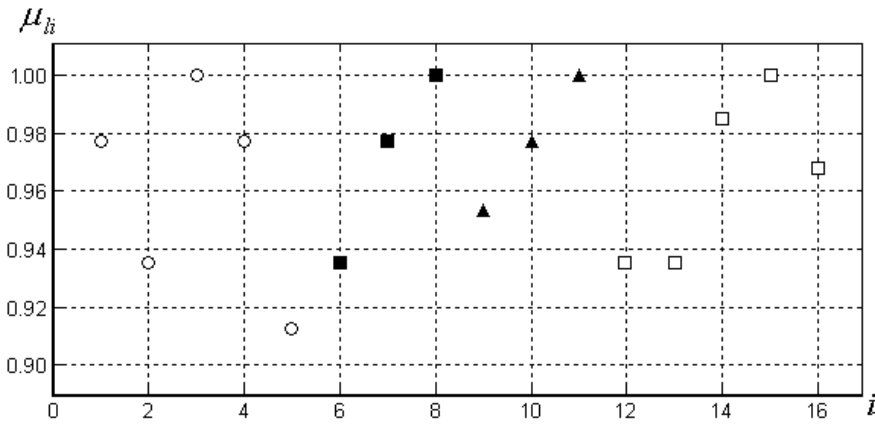
Figure 10. Membership functions of four fully separated fuzzy clusters

The third object is the typical point of the first fuzzy cluster, the eighth object is the typical point of the second fuzzy cluster, the eleventh object is the typical point of the third fuzzy cluster, and the fifteenth object is the typical point of the corresponding cluster. Main characteristics of fuzzy clusters are presented in Table 2.

Table 2. Main characteristics of fuzzy clusters

| Number of fuzzy clusters	Characteristics of fuzzy clusters in the obtained allotment		
	Value of the index (34)	Value of the index (37)	Value of the index (39)
Class 1	0.07743056	0.09997046	0.96128472
Class 2	0.05671297	0.07734306	0.97164352
Class 3	0.04456019	0.05783274	0.97771991
Class 4	0.06875000	0.08579271	0.96562500

4.3. The Anderson's Iris data set

Let us consider an application of proposed technique to the classification problem for the Anderson's Iris data set [1]. The Anderson's Iris data is the most known database to be found in the pattern recognition literature. The data set represents different categories of iris plants having four attribute values. The four attribute values represent the sepal length, sepal width, petal length and petal width measured for 150 irises. It has three classes Setosa, Versicolor and Virginica, with 50 samples per class.

The problem is to classify the plants into three subspecies on the basis of this information. It is known that two classes Versicolor and Virginica have some amount of overlap while the class Setosa is linearly separable from the other two. The real assignments to the three classes are presented in Table 3.

**Table 3.   Real object assignments in the Anderson's Iris data set**

Class		Numbers of objects
Number	Name	
1	SETOSA	1, 6, 10, 18, 26, 31, 36, 37, 40, 42, 44, 47, 50, 51, 53, 54, 55, 58, 59, 60, 63, 64, 67, 68, 71, 72, 78, 79, 87, 88, 91, 95, 96, 100, 101, 106, 107, 112, 115, 124, 125, 134, 135, 136, 138, 139, 143, 144, 145, 149
2	VERSICOLOR	3, 8, 9, 11, 12, 14, 19, 22, 28, 29, 30, 33, 38, 43, 48, 61, 65, 66, 69, 70, 76, 84, 85, 86, 92, 93, 94, 97, 98, 99, 103, 105, 109, 113, 114, 116, 117, 118, 119, 120, 121, 128, 129, 130, 133, 140, 141, 142, 147, 150
3	VIRGINICA	2, 4, 5, 7, 13, 15, 16, 17, 20, 21, 23, 24, 25, 27, 32, 34, 35, 39, 41, 45, 46, 49, 52, 56, 57, 62, 73, 74, 75, 77, 80, 81, 82, 83, 89, 90, 102, 104, 108, 110, 111, 122, 123, 126, 127, 131, 132, 137, 146, 148

Let us consider the effectiveness of the proposed technique by testing the Anderson's Iris data set. The initial data set was preprocessed by using formulae (28), (32), (33) and the matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, 150$ was constructed. The technique for detecting the optimal number of elements in fuzzy clusters was also applied to the matrix of fuzzy tolerance for $u \in [u_{\min} = 51, u_{\max} = 54]$ using validity measures (41) – (43). The performance of validity measures is shown in Figures 11 – 13.
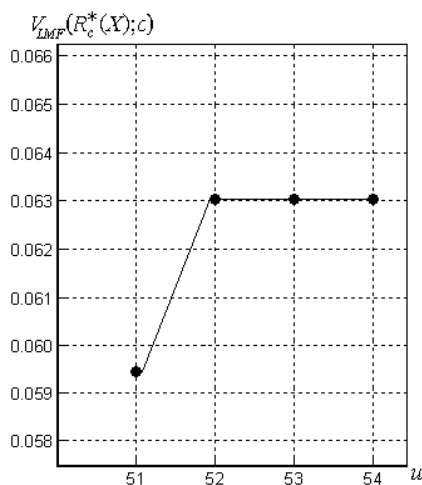
Figure 11.   Plot of the linear measure of fuzziness for the Anderson's data set

$V_{QMF}(R_c^*(X);c)$
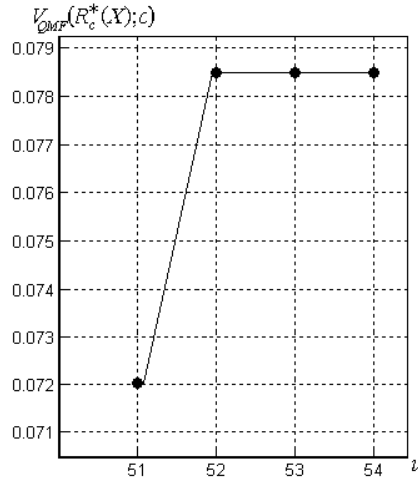
Figure 12. Plot of the quadratic measure of fuzziness for the Anderson's data set


$V_{MSC}(R_c^*(X);c)$
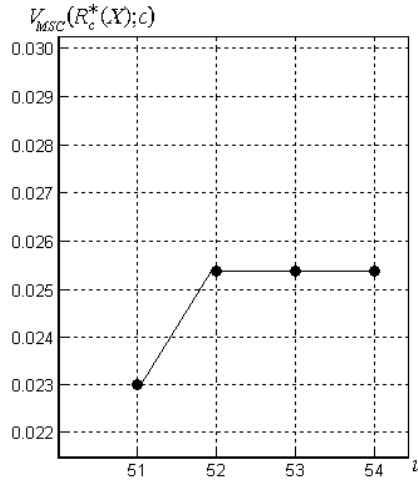
Figure 13. Plot of the measure of separation and compactness for the Anderson's data set

By executing the proposed technique, the allotment $R_c^*(X)$ among three fuzzy clusters, which corresponds to the result, is received for the value $u = 52$. The allotment was obtained for the tolerance threshold $\alpha = 0.964067$. It is should be noted, that the actual number of classes equal 3 and this number corresponds to the number of fuzzy clusters in the obtained allotment $R_c^*(X)$.

Membership functions of three classes are presented in Figure 14, where membership values of the first class are represented by +, membership values of the second class are

represented by ■, and membership values of the third class are represented by ○. Values which equal zero are not shown in Figure 14.
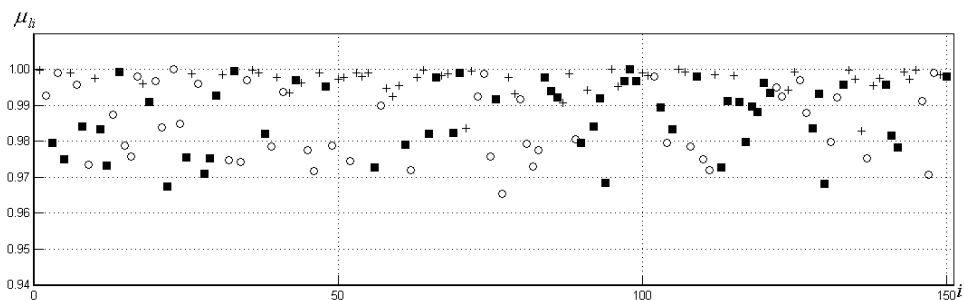
Figure 14. Membership functions of three fully separated fuzzy clusters

Thus, we obtain the following: the first class is formed by 50 elements all being Iris Setosa; the second class is formed by 52 elements, 48 of them being Iris Versicolor and 4 Iris Virginica; the third class by 48 elements, 46 of them being Iris Virginica and 2 Iris Versicolor. So, there are six mistakes of classifications. The object assignments resulting from the proposed technique application to the Anderson's Iris data are presented in Table 4. Misclassified objects are bolded in Table 4.

Table 4. The results of the proposed technique application: the object assignments

Class		Numbers of objects
Number	Name	
1	SETOSA	1, 6, 10, 18, 26, 31, 36, 37, 40, 42, 44, 47, 50, 51, 53, 54, 55, 58, 59, 60, 63, 64, 67, 68, 71, 72, 78, 79, 87, 88, 91, 95, 96, 100, 101, 106, 107, 112, 115, 124, 125, 134, 135, 136, 138, 139, 143, 144, 145, 149
2	VERSICOLOR	3, **5**, 8, 11, 12, 14, 19, 22, **25**, 28, 29, 30, 33, 38, 43, 48, **56**, 61, 65, 66, 69, 70, 76, 84, 85, 86, **90**, 92, 93, 94, 97, 98, 99, 103, 105, 109, 113, 114, 116, 117, 118, 119, 120, 121, 128, 129, 130, 133, 140, 141, 142, 150
3	VIRGINICA	2, 4, 7, **9**, 13, 15, 16, 17, 20, 21, 23, 24, 27, 32, 34, 35, 39, 41, 45, 46, 49, 52, 57, 62, 73, 74, 75, 77, 80, 81, 82, 83, 89, 102, 104, 108, 110, 111, 122, 123, 126, 127, 131, 132, 137, 146, **147**, 148

The value of the membership function of the fuzzy cluster which corresponds to the first class is maximal for the ninety-fifth object and is equal one. That is why the ninety-fifth object is the typical point of the fuzzy cluster which corresponds to the Setosa class. The membership value of the ninety-eighth object is equal one for the fuzzy cluster which corresponds to the second class. Thus, the ninety-eighth object is the typical point of the fuzzy cluster which corresponds to the Versicolor class. The membership value of the

twenty-third object is equal one for the fuzzy cluster which corresponds to the third class. So, the twenty-third object is the typical point of the fuzzy cluster which corresponds to the Virginica class.

In order to compare the proposed technique with other clustering procedures, the D-AFC(c)-algorithm was applied to the matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1,\ldots,150$ with the number of classes $c = 3$. By executing the D-AFC(c)-algorithm, the allotment $R_c^*(X)$ among three fuzzy clusters, which corresponds to the result, is received for the tolerance threshold $\alpha = 0.964149$. Membership functions of three classes are presented in Figure 15.
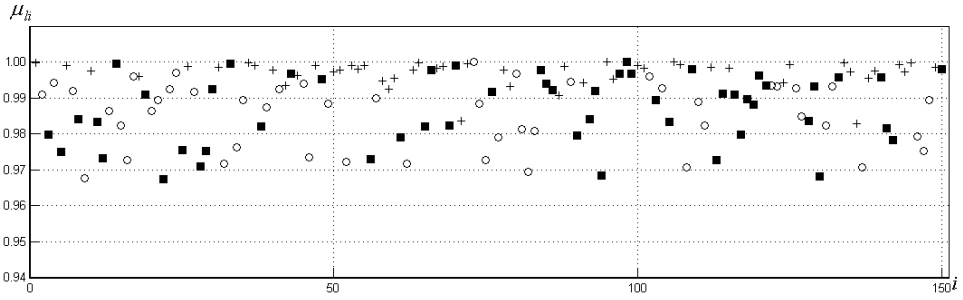
Figure 15. Membership functions of three fully separated fuzzy clusters obtained from the D-AFC(c)-algorithm

So, the object assignment obtained from the D-AFC(c)-algorithm is equal to the object assignment obtained from the proposed technique. However, membership functions of fuzzy clusters obtained from the D-AFC(c)-algorithm are differ from membership functions of fuzzy clusters obtained from the proposed technique. In particular, the ninety-fifth object is the typical point of the fuzzy cluster which corresponds to the Setosa class, the ninety-eighth object is the typical point of the fuzzy cluster which corresponds to the Versicolor class, and the seventy-third object is the typical point of the fuzzy cluster which corresponds to the Virginica class.

5. Conclusions

Results of experiments are summarized and discussed in the first subsection of the section. The second subsection deals with the perspectives on future investigations.

5.1. Discussion

The technique for detecting the optimal maximal number of elements in fuzzy clusters is proposed in the paper. The direct relational D-AFC(u)-algorithm of possibilistic clustering

and some validity measure are the basis of the proposed technique. The essence of the proposed technique is setting the relationship between the optimal maximal number of elements in each fuzzy cluster, and the number of fuzzy clusters in the sought clustering structure. The allotment among a priori unknown number of fully separate fuzzy clusters with optimal maximal number of elements in the each fuzzy cluster is the result of the data processing.

The results, obtained in the first two experiments by using the proposed technique, are equal to the results, obtained from the direct relational D-PAFC-algorithm of possibilistic clustering [22]. On the other hand, the result, obtained by using the proposed technique for the Anderson's Iris data, is differing from the results, obtained from the D-PAFC-algorithm and D-AFC(c)-algorithm of possibilistic clustering.

Thus, the results of application of the proposed technique to three well-known data sets 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 clusters.

5.2. Perspectives

Possibility degrees for values of the most possible number of fuzzy clusters in the sought allotment can be calculated [22]. So, the relationship between possibility degrees for values of the number of fuzzy clusters and values of the number of elements in fuzzy clusters should be investigated.

The perspective for investigations is of great interest both from the theoretical point of view and from the practical one as well.

Acknowledgments

References

[1] Anderson E., The irises of the Gaspe Peninsula, *Bulletin of the American Iris Society*, **59**, 1, 1935, 2-5.
[2] Bezdek J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.
[3] Chiang J.-H., Yue S., Yin Z.-X., A new fuzzy cover approach to clustering, *IEEE Transactions on Fuzzy Systems*, **12**, 2, 2004, 199-208.
[4] Corsini P., Lazzerini B., Marcelloni F., A new fuzzy relational clustering algorithm based on the fuzzy C-means algorithm, *Soft Computing*, **9**, 6, 2005, 439-447.

[5] De Cáceres M., Oliva F., Font X., On relational possibilistic clustering, *Pattern Recognition*, **39**, 11, 2006, 2010-2024.

[6] Everitt B.S., Landau S., Leese M., Stahl D., *Cluster Analysis, 5th Edition*, Wiley, Chichester, 2011.

[7] Hamasuna Y., Endo Y., Miyamoto S., Fuzzy *C*-means clustering for data with clusterwise tolerance based on L_2- and L_1-regularization, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **15**, 1, 2011, 68-75.

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

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

[10] Komazaki Y., Miyamoto S., Variables for controlling cluster sizes on fuzzy *C*-means, in: V. Torra, Y. Narukawa, G. Navarro-Arribas, D. Megías (eds.), *Modeling Decisions for Artificial Intelligence: Proceedings of the 10th International Conference MDAI'2013, Barcelona, Spain, November 20-22, 2013*, Springer, Berlin, 2013, 192-203.

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

[12] Łęski J.M., Robust possibilistic clustering, *Archives of Control Sciences*, **10**, 3/4, 2000, 141-155.

[13] Mandel I.D., *Clustering Analysis*, Finansy i Statistica, Moscow, 1988. (in Russian)

[14] Ménard M., Courboulay V., Dardignac P.-A., Possibilistic and probabilistic fuzzy clustering: unification within the framework of the non-extensive thermostatistics, *Pattern Recognition*, **36**, 6, 2003, 1325-1342.

[15] Miyamoto S., Ichihashi H., Honda K., *Algorithms for Fuzzy Clustering. Methods in C-Means Clustering with Applications*, Springer, Berlin, 2008.

[16] Miyamoto S., Different objective functions in fuzzy C-means algorithms and kernel-based clustering, *International Journal of Fuzzy Systems*, **13**, 2, 2011, 89-97.

[17] Pedrycz W., Fuzzy sets in pattern recognition: methodology and methods, *Pattern Recognition*, **23**, 1/2, 1990, 121-146.

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

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

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

[21] Vapnik V.N., *Statistical Learning Theory*, Wiley, New York, 1998.

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

[23] Viattchenin D.A., Damaratski A., Direct heuristic algorithms of possibilistic clustering based on transitive approximation of fuzzy tolerance, *Informatica Economicá Journal*, **17**, 3, 2013, 5-15.

[24] Viattchenin D.A., Yaroma A., Damaratski A., A novel direct relational heuristic algorithm of possibilistic clustering, *International Journal of Computer Applications*, **107**, 18, 2014, 15-21.

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

[26] Xie Z., Wang S.T., Chung F.L., An enhanced possibilistic c-means clustering algorithm EPCM, *Soft Computing*, **12**, 6, 2008, 593-611.

[27] Yang M.-S., Wu K.-L., Unsupervised possibilistic clustering, *Pattern Recognition*, **39**, 1, 2006, 5-21.