Direct Heuristic Algorithms of Possibilistic Clustering Based on Transitive Approximation of Fuzzy Tolerance

Dmitri A. VIATTCHENIN, Aliaksandr DAMARATSKI
United Institute of Informatics Problems of the National Academy of Sciences of Belarus
Surganov St. 6, 220012 Minsk, Belarus
viattchenin@mail.ru, a.damaratski@gmail.com

This paper deals with the problem of a heuristic approach to possibilistic clustering. The approach is based on the concept of allotment among fuzzy clusters. The paper provides the description of basic concepts of the heuristic approach to possibilistic clustering. Plans of direct prototype-based heuristic algorithms of possibilistic clustering based on a transitive approximation of a fuzzy tolerance are described in detail. An illustrative example of application of the basic version of the proposed algorithms to Sneath and Sokal's two-dimensional data set is considered. Preliminary conclusions are formulated.
Keywords: *Clustering, Fuzzy Tolerance, Transitive Approximation, Allotment, Fuzzy Cluster*

1 Introduction

Some remarks on heuristic possibilistic clustering are considered in the first subsection of the section. The second subsection includes notes about fuzzy approximations of a fuzzy tolerance.

1.1 Preliminary Remarks

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

Clustering algorithms in general can also be divided into two types: hard versus fuzzy. Hard clustering assigns each object to exactly one cluster. In fuzzy clustering, a given pattern does not necessarily belong to only one cluster, but can have varying degrees of memberships to several clusters. Heuristic methods of fuzzy clustering, hierarchical methods of fuzzy clustering and optimization methods of fuzzy clustering were proposed by different researchers. The most widespread approach in fuzzy clustering is the optimization approach. Moreover, a possibilistic approach to clustering was proposed by Krishnapuram and Keller [1] and the possibilistic approach to clustering can be considered as a way in the optimization approach in fuzzy clustering because all methods of possibilistic clustering are objective function-based methods.

On the other hand, a heuristic approach to possibilistic clustering was proposed in [2]. The essence of the proposed 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 pairwise similarity of objects. Heuristic algorithms of possibilistic clustering display high level of essential clarity and low level of a complexity.

Heuristic clustering algorithms which are based on a definition of the cluster concept are called algorithms of direct classification or direct clustering algorithms. 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 family of direct prototype-based heuristic algorithms of possibilistic clustering includes

- D-AFC-TC-algorithm: using the construc-

tion 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 all direct prototype-based heuristic possibilistic clustering algorithms are based on a transitive closure of an initial fuzzy tolerance relation.

1.2 On a Transitive Approximation of a Fuzzy Tolerance

Let $X = \{x_1,..., x_n\}$ be the initial set of elements and $T : X \times X \to [0,1]$ some binary fuzzy relation on X with $\mu_T(x_i, x_j) \in [0,1]$, $\forall x_i, x_j \in X$ being its membership function. A square matrix $T = [\mu_T(x_i, x_j)]$ of size n with elements $\mu_T(x_i, x_j)$ in the unit interval $[0,1]$ is called a fuzzy tolerance matrix if it is reflexive and symmetric. In other words, conditions $\mu_T(x_i, x_i) = 1$, $\forall i \in \{1,...,n\}$ and $\mu_T(x_i, x_j) = \mu_T(x_j, x_i)$, $\forall(i, j) \in \{1,...,n\}^2$ are met. If a condition of min-transitivity $\min(\mu_T(x_i, x_j), \mu_T(x_j, x_k)) \leq \mu_T(x_i, x_k)$, $\forall(i, j, k) \in \{1,...,n\}^3$ is met for the matrix, then the matrix is called a fuzzy equivalence matrix.

Fuzzy tolerance matrices do not naturally possess property of transitivity. So, transitivity needs to be imposed artificially into fuzzy tolerance matrices. Computing a unique transitive closure of a fuzzy tolerance matrix is the possibilities for turning a fuzzy tolerance matrix into fuzzy equivalence matrix. A method for construction of the transitive closure of a fuzzy tolerance matrix is given, for example, in [3]. Other possibilities are the computation of a transitive opening of a fuzzy tolerance matrix, where allowed that elements are either raised or lowered with respect to their initial value [4]. A transitive opening of a fuzzy tolerance is the reflexive,

symmetric and min-transitive fuzzy relation. Notice that it can be several transitive openings of a fuzzy tolerance. An effective algorithm to computing the maximal transitive opening of a fuzzy tolerance is proposed in [4].

Computing a transitive approximation of a fuzzy tolerance is more general possibility for turning for turning a fuzzy tolerance matrix into fuzzy equivalence matrix than computing a transitive closure or a transitive opening of a fuzzy tolerance. The corresponding TAGA-algorithm is proposed and investigated in [5]. Different approximations of an initial fuzzy tolerance matrix can be generated depending on the selection of an aggregation operator $f(y_1, y_2,..., y_m)$ that plays a key role in the TAGA-algorithm. It is assumed that the arguments $y_1,..., y_m$ of f appear in increasing order of magnitude:

1) maximum: $f_1(y_1, y_2,..., y_m) = \max_i(y_i) = y_m$;

2) minimum: $f_2(y_1, y_2,..., y_m) = \min_i(y_i) = y_1$;

3) mean: $f_3(y_1, y_2,..., y_m) = \frac{1}{m}\sum_i y_i$;

4) median:
$$f_4(y_1, y_2,..., y_m) = \begin{cases} y_{(m+1)/2}, & \text{if } m \text{ is odd} \\ \dfrac{(y_{m/2} + y_{m/2+1})}{2}, & \text{if } m \text{ is even} \end{cases}$$

5) upmedian: $f_5(y_1, y_2,..., y_m) = y_{\lceil(m+1)/2\rceil}$;

6) downmedian: $f_6(y_1, y_2,..., y_m) = y_{\lfloor(m+1)/2\rfloor}$.

So, a transitive approximation $\tilde{T}_k$, $k \in \{1,...,6\}$ shall be constructed by using the TAGA-algorithm according to the corresponding aggregation operator $f_k(y_1, y_2,..., y_m)$, $k \in \{1,...,6\}$. Note that the approximation $\tilde{T}_1$ which is generated by the maximum operator $f_1(y_1, y_2,..., y_m)$ is the transitive closure of T.

The TAGA-algorithm can be inserted into direct prototype-based heuristic algorithms of possibilistic clustering instead the procedure of computing a transitive closure of a fuzzy tolerance. So, the main goal of this paper is a consideration of new direct prototype-based heuristic algorithms of possibilistic clustering based on a transitive approximation of a fuzzy tolerance. The contents of this paper is as follows: in the second section basic con-

cepts of the heuristic approach to possibilistic clustering are considered and plans of new algorithms are described, in the third section an illustrative example of application of the basic version of proposed new algorithms to Sneath and Sokal's [6] artificial data set is given, in the fourth section some final remarks are stated.

2 Content Details

Basic definitions of a heuristic approach to possibilistic clustering are considered in the first subsection of the section. The second subsection includes remarks on the data pre-processing. A leap heuristic and a procedure for automatic selection of an aggregate operator are given in the third subsection. Plans of proposed algorithms are described in the fourth subsection.

2.1 Basic Concepts of the Clustering Method

Let us remind the basic concepts of the heuristic method of possibilistic clustering. 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

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

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 cluster $A^l_{(\alpha)}$. So, condition

$A^l_\alpha = Supp(A^l_{(\alpha)})$ is met for each fuzzy cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1, n]$. Membership degree can be interpreted as a degree of typicality of an element to a fuzzy cluster.

Let T is a fuzzy tolerance on X, where X is the set of objects, and $\{A^1_{(\alpha)}, ..., A^n_{(\alpha)}\}$ is the family of fuzzy clusters for some $\alpha \in (0,1]$. The point $\tau^l_e \in A^l_\alpha$, for which

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

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

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

is met for all fuzzy 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 $X = \{x_1, ..., x_n\}$ among fuzzy 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.

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 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 $A^l_{(\alpha)}$, $l = \overline{1,c}$, $\alpha \in (0,1]$ 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 $R^\alpha_{c(z)}(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, c \leq n, \alpha \in (0,1]\}$ corresponds to the formulation of a concrete prob-

lem, then this allotment is an adequate allotment. In particular, if a condition

$$\bigcup_{l=1}^{c} A_\alpha^l = X , \qquad (4)$$

and a condition

$$card(A_\alpha^l \cap A_\alpha^m) = 0 , \ \forall A_{(\alpha)}^l, A_{(\alpha)}^m , \ l \neq m , \ \alpha \in (0,1], \ (5)$$

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

Allotment $R_P^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}\}$ of the set of objects among the minimal number c, $2 \leq c \leq n$ of fully separate fuzzy 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. The criterion

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

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 (6) 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

$$R_c^*(X) = \arg \max_{R_{c(z)}^\alpha(X) \in B} F(R_{c(z)}^\alpha(X), \alpha) . \qquad (7)$$

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.

2.2 On the Data Preprocessing

The initial data should be presented as a matrix of attributes $\hat{X}_{n \times m} = [\hat{x}_i^t]$, $i = 1,...,n$, $t = 1,...,m$, where the value $\hat{x}_i^t$ is the value of the t-th attribute for i-th object. Thus, the proposed approach to clustering can be used with the data matrix by choosing a suitable metric to measure similarity. The two-way data can be normalized, for example, as follows:

$$x_i^t = \frac{\hat{x}_i^t}{\max\limits_i \hat{x}_i^t} , \qquad (8)$$

or using a formula

$$x_i^t = \frac{\hat{x}_i^t - \min\limits_i \hat{x}_i^t}{\max\limits_i \hat{x}_i^t - \min\limits_i \hat{x}_i^t} \qquad (9)$$

So, each object can be considered as a fuzzy set x_i, $i = 1,...,n$ and $x_i^t = \mu_{x_i}(x^t) \in [0,1]$, $i = 1,...,n$, $t = 1,...,m$ are their membership functions. Of course, some other methods for the data normalization are described in different bibliographical sources.

The matrix of coefficients of pair wise dissimilarity between objects $I = [\mu_I(x_i, x_j)]$, $i,j = 1,...,n$ can be obtained after application of some distance function to the matrix of normalized data $X_{n \times m} = [\mu_{x_i}(x^t)]$, $i = 1,...,n$, $t = 1,...,m$. In particular, the squared normalized Euclidean distance

$$d(x_i, x_j) = \frac{1}{m} \sum_{t=1}^{m} \left(\mu_{x_i}(x^t) - \mu_{x_j}(x^t) \right)^2, \ i,j = 1,...,n, \ (10)$$

is the widely used distance for fuzzy sets x_i, x_j, $i,j = 1,...,n$ in $X = \{x_1,..., x_n\}$. The squared normalized Euclidean distance (10) does not satisfy the min-transitivity condition. The fact was demonstrated in [3] where some other distances for fuzzy sets are also considered. The matrix of fuzzy tolerance $I = [\mu_I(x_i, x_j)]$, $i,j = 1,...,n$ can be obtained after application of complement operation to the matrix of fuzzy intolerance $I = [\mu_I(x_i, x_j)]$. A distance $d(x_i, x_j)$ for fuzzy sets is a general parameter of direct prototype-based heuristic algorithms of possibilistic clustering.

2.3 Some Additional Remarks

The unique allotment among unknown number c of fuzzy clusters can be selected from the set of allotments which depends on the tolerance threshold. An idea of a leap in similarity values for finding of the appropriate value α_ℓ of the tolerance threshold can be useful for the aim. That is why some appropriate value α_ℓ of the tolerance threshold must be detected in the ordered sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \le 1$. For the purpose, a leap heuristic can be used and the heuristic can be described as a three-step procedure [2]:

1. Values $g_\ell = \alpha_{\ell-1}/\alpha_\ell$ should be computed for all α_ℓ, $\ell = 1,\ldots,Z-1$ in the ordered sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \le 1$;

2. The value α_{ℓ^*} which corresponds to ℓ^* for some $g_{\ell^*} = \min_\ell g_\ell$, $\ell = 1,\ldots,Z-1$ should be selected;

3. If a few minimal g_{ℓ^*} values are obtained in the set $\{g_\ell\}$, $\ell = 1,\ldots,Z-1$, then the value α_{ℓ^*}, which corresponds to the minimal value of the index ℓ, should be selected.

An aggregation operator $f_k(y_1, y_2, \ldots, y_m)$, $k \in \{1,\ldots,6\}$ is the parameter of the TAGA-algorithm and the parameter can be selected either by operator or automatically. Automatic choosing of the aggregation operator can be made on a basis of estimation of the corresponding transitive approximations $\tilde{T}_k$, $k \in \{1,\ldots,6\}$. In particular, a distance between fuzzy relations can be used for the purpose. The distance between fuzzy relations is given in [2]. For a case of a fuzzy tolerance T and its transitive approximation $\tilde{T}_k$, $k \in \{1,\ldots,6\}$, the distance can be written as follows

$$d(T,\tilde{T}_k) = \sum_{(x_i,x_j)} \left| \mu_T(x_i,x_j) - \mu_{\tilde{T}_k}(x_i,x_j) \right|,$$

$$\tilde{T}_k \in \{\tilde{T}_1,\ldots,\tilde{T}_6\}. \qquad (11)$$

A preferable fuzzy relation for clustering procedure can be selected from the set $\{\tilde{T}_1, \tilde{T}_2, \tilde{T}_3, \tilde{T}_4, \tilde{T}_5, \tilde{T}_6\}$ according to the condition $\min_k d(T,\tilde{T}_k)$, $k = 1,\ldots 6$. So, constructing an acceptable transitive approximation $\tilde{T}_k$,

$k \in \{1,\ldots,6\}$ of the initial fuzzy tolerance T can be described as a two-step procedure:

1. Check the following condition:
   if the aggregation operator $f_k(y_1, y_2, \ldots, y_m)$, $k \in \{1,\ldots,6\}$ is selected by expert
   then the corresponding transitive approximation $\tilde{T}_k$, $k \in \{1,\ldots,6\}$ should be constructed by using the TAGA-algorithm according to the selected aggregation operator and stop
   else go to step 2;

2. Perform the following operations:

2.1 Construct the set $\{\tilde{T}_1,\ldots,\tilde{T}_6\}$ of transitive approximations of the initial fuzzy tolerance T by applying the TAGA-algorithm to the fuzzy tolerance T according to each aggregation operator $f_k(y_1, y_2, \ldots, y_m)$, $k \in \{1,\ldots,6\}$;

2.2 Calculate the value of distance (11) from the fuzzy tolerance T to each fuzzy relation from the set $\{\tilde{T}_1,\ldots,\tilde{T}_6\}$;

2.3 Check the following condition:
   if for some fuzzy relation $\tilde{T}_k \in \{\tilde{T}_1,\ldots,\tilde{T}_6\}$ the condition $\min_k d(T,\tilde{T}_k)$, $k = 1,\ldots 6$ is met
   then the corresponding transitive approximation $\tilde{T}_k$, $k \in \{1,\ldots,6\}$ is the most preferable fuzzy relation for clustering and stop.

So, a transitive approximation of the fuzzy tolerance which is obtained in step 2 of the procedure will be most near to the fuzzy tolerance in sense of the distance (11).

2.4 Plans of Clustering Procedures

Detection of the unique allotment $R_c^*(X)$ among a priori unknown number c of fully separated fuzzy clusters is the matter of proposed algorithms. The D-AFC-TAGA-algorithm is a nine-step procedure of classification:

1. Construct the matrix of the fuzzy tolerance relation $T_{n \times n} = [\mu_T(x_i,x_j)]$ by normalizing the initial data $\hat{X}_{n \times m} = [\hat{x}_i^t]$, $i = 1,\ldots,n$, $t = 1,\ldots,m$, and by choosing a suitable distance for fuzzy sets;

2. Construct an acceptable transitive approximation $\tilde{T}_k$ of the fuzzy tolerance T;

3. Construct the ordered sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \leq 1$ of α-levels for the transitive approximation $\tilde{T}_k$ of the fuzzy tolerance T; the value α_ℓ must be found using the leap heuristic;

4. Construct the fuzzy relation $\tilde{T}_{k(\alpha_\ell)}$ for the value α_ℓ;

5. Construct the initial allotment $R_I^\alpha(X) = \{A_{(\alpha_\ell)}^l\}$ for the fuzzy relation $\tilde{T}_{k(\alpha_\ell)}$; construct the allotments, which satisfy conditions (4) and (5);

6. Construct the class of possible solutions of the classification problem $B(\alpha_\ell) = \{R_{c(z)}^{\alpha_\ell}(X)\}$ and calculate the value of criterion (6) for every allotment $R_{c(z)}^{\alpha_\ell}(X) \in B(\alpha_\ell)$;

7. Check the following condition:

   if for some unique allotment $R_{c(z)}^{\alpha_\ell}(X) \in B(\alpha_\ell)$ the condition (7) is met

   then this allotment is the classification result $R_c^*(X)$ for the value α_ℓ and stop

   else construct the set of allotments $B'(\alpha_\ell) \subseteq B(\alpha_\ell)$ which satisfy condition (7) and go to step 8;

8. Perform the following operations for each allotment $R_{c(z)}^{\alpha_\ell}(X) \in B'(\alpha_\ell)$:

   8.1   Let $l := 1$;

   8.2   Find the support $Supp(A_{(\alpha_\ell)}^l) = A_{\alpha_\ell}^l$ of the fuzzy cluster $A_{(\alpha_\ell)}^l \in R_{c(z)}^{\alpha_\ell}(X)$ and construct the matrix of attributes $X_{n_l \times m} = [x_i^t]$, $x_i \in A_{\alpha_\ell}^l$, $t = 1, \ldots, m$ for $A_{\alpha_\ell}^l$ where $n_l = card(A_{\alpha_\ell}^l)$;

   8.3   Calculate the prototype $\tau^l = \{\bar{x}^1, \ldots \bar{x}^m\}$ of the class $A_{\alpha_\ell}^l$ according to the formula $\bar{x}^t = \dfrac{1}{n_l} \sum_{x_i \in A_{\alpha_\ell}^l} x_i^t$, $t = 1, \ldots, m$;

   8.4   Calculate the distance $d(\tau^l, \tau^l)$ between the typical point τ^l of the fuzzy cluster $A_{(\alpha_\ell)}^l$ and its prototype τ^l;

   8.5   Check the following condition:

   if all fuzzy clusters $A_{(\alpha_\ell)}^l \in R_{c(z)}^{\alpha_\ell}(X)$ are not verified

   then let $l = l + 1$ and go to step 8.2
   else go to step 9

9. Compare the fuzzy clusters $A_{(\alpha_\ell)}^l$ which are elements of different allotments $R_{c(z)}^{\alpha_\ell}(X) \in B'(\alpha_\ell)$; the allotment $R_{c(z)}^{\alpha_\ell}(X) \in B'(\alpha_\ell)$ for which the distance $d(\tau^l, \tau^l)$ is minimal for all fuzzy clusters $A_{(\alpha_\ell)}^l$ is the classification result $R_c^*(X)$.

The allotment $R_c^*(X) = \{A_{(\alpha)}^l \mid l = \overline{1,c}\}$ among the unknown number c of fully separate fuzzy clusters, normalized prototypes $\{\tau^1, \ldots, \tau^c\}$ of the corresponding fuzzy clusters, and the value of tolerance threshold $\alpha \in (0,1]$ are results of classification. The D-AFC-TAGA-algorithm is a basic version of the family of proposed clustering procedures based on computing a transitive approximation of the initial fuzzy tolerance.

Detection of an unknown minimal number c of compact and well-separated fuzzy clusters can be considered as the aim of classification in some situations. So, the D-PAFC-TAGA-algorithm is a nine-step procedure of classification:

1. Construct the matrix of the fuzzy tolerance relation $T_{n \times n} = [\mu_T(x_i, x_j)]$ by normalizing the initial data $\hat{X}_{n \times m} = [\hat{x}_i^t]$, $i = 1, \ldots, n$, $t = 1, \ldots, m$, and by choosing a suitable distance for fuzzy sets;

2. Construct an acceptable transitive approximation $\tilde{T}_k$ of the fuzzy tolerance T;

3. Construct the ordered sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \leq 1$ of α-levels for the transitive approximation $\tilde{T}_k$ of the fuzzy tolerance T;

4. Construct the fuzzy relation $\tilde{T}_{k(\alpha_1)}$ for the value α_1;

5. Construct the initial allotment $R_I^\alpha(X) = \{A_{(\alpha_1)}^l\}$ for the fuzzy relation $\tilde{T}_{k(\alpha_1)}$; construct the allotments, which satisfy conditions (4) and (5);

6. Construct the class of possible solutions of the classification problem $B(\alpha_1) = \{R_{c(z)}^{\alpha_1}(X)\}$ and calculate the value of

criterion (6) for every allotment $R_{c(z)}^{\alpha_1}(X) \in B(\alpha_1)$;

7. Check the following condition:

    if for some unique allotment $R_{c(z)}^{\alpha_1}(X) \in B(\alpha_1)$ the condition (7) is met

    then this allotment is the classification result $R_P^{\alpha_1}(X)$ for the value α_1 and stop

    else construct the set of allotments $B'(\alpha_1) \subseteq B(\alpha_1)$ which satisfy condition (7) and go to step 8;

8. Perform the following operations for each allotment $R_{c(z)}^{\alpha_1}(X) \in B'(\alpha_1)$:

    8.1   Let $l = 1$;

    8.2   Find the support $Supp(A_{(\alpha_1)}^l) = A_{\alpha_1}^l$ of the fuzzy cluster $A_{(\alpha_1)}^l \in R_{c(z)}^{\alpha_1}(X)$ and construct the matrix of attributes $X_{n_l \times m} = [x_i^t]$, $x_i \in A_{\alpha_1}^l$, $t = 1, \ldots, m$ for $A_{\alpha_1}^l$ where $n_l = card(A_{\alpha_1}^l)$;

    8.3   Calculate the prototype $\tau^l = \{x^1, \ldots, x^m\}$ of the class $A_{\alpha_1}^l$ according to the formula $\bar{x}^t = \frac{1}{n_l} \sum_{x_i \in A_{\alpha_1}^l} x_i^t$, $t = 1, \ldots, m$;

    8.4   Calculate the distance $d(\tau^l, \tau^l)$ between the typical point τ^l of the fuzzy cluster $A_{\alpha_1}^l$ and its prototype τ^l;

    8.5   Check the following condition:

    if all fuzzy clusters $A_{(\alpha_1)}^l \in R_{c(z)}^{\alpha_1}(X)$ are not verified

    then let $l = l + 1$ and go to step 8.2
    else go to step 9

9. Compare the fuzzy clusters $A_{\alpha_1}^l$ which are elements of different allotments $R_{c(z)}^{\alpha_1}(X) \in B'(\alpha_1)$, and the allotment $R_{c(z)}^{\alpha_1}(X) \in B'(\alpha_1)$ for which the distance $d(\tau^l, \tau^l)$ is minimal for all fuzzy clusters $A_{\alpha_1}^l$ is the classification result $R_P^{\alpha_1}(X)$.

The leap heuristic is not used in the D-PAFC-TAGA-algorithm because the solution of the classification problem should be obtained for the value α_1. The principal allotment $R_P^{\alpha_1}(X) = \{A_{(\alpha_1)}^l \mid l = \overline{1,c}\}$ among the minimal number c of fully separate fuzzy clus-

ters, normalized prototypes $\{\tau^1, \ldots, \tau^c\}$ of the corresponding fuzzy clusters, and the value of tolerance threshold $\alpha_1 \in (0,1]$ are results of classification.

The determination of the allotment $R_c^*(X)$ among an unknown number c of fully separated fuzzy clusters with respect to the minimal value α of the tolerance threshold is the essence of the D-AFC-TAGA(α)-algorithm which can be summarized as a ten-step classification procedure:

1. Construct the matrix of the fuzzy tolerance relation $T_{n \times n} = [\mu_T(x_i, x_j)]$ by normalizing the initial data $\hat{X}_{n \times m} = [\hat{x}_i^t]$, $i = 1, \ldots, n$, $t = 1, \ldots, m$, and by choosing a suitable distance for fuzzy sets;

2. Construct an acceptable transitive approximation $\tilde{T}_k$ of the fuzzy tolerance T;

3. Construct the ordered sequence $0 < \alpha_0 < \alpha_1 < \ldots < \alpha_\ell < \ldots < \alpha_Z \leq 1$ of α-levels for the transitive approximation $\tilde{T}_k$ of the fuzzy tolerance T; the value α_ℓ must be found using the leap heuristic;

4. Check the following condition:

    if for the calculated value α_ℓ the condition $\alpha \leq \alpha_\ell$ is met
    then go to step 5
    else the value α_ℓ must be found by the leap heuristic in the ordered sequence $\alpha^* < \ldots < \alpha_\ell < \ldots < \alpha_Z \leq 1$ and go to step 5;

5. Construct the fuzzy relation $\tilde{T}_{k(\alpha_\ell)}$ for the value α_ℓ;

6. Construct the initial allotment $R_I^\alpha(X) = \{A_{(\alpha_\ell)}^l\}$ for the fuzzy relation $\tilde{T}_{k(\alpha_\ell)}$; construct the allotments, which satisfy conditions (4) and (5);

7. Construct the class of possible solutions of the classification problem $B(\alpha_\ell) = \{R_{c(z)}^{\alpha_\ell}(X)\}$ and calculate the value of criterion (6) for every allotment $R_{c(z)}^{\alpha_\ell}(X) \in B(\alpha_\ell)$;

8. Check the following condition:

    if for some unique allotment $R_{c(z)}^{\alpha_\ell}(X) \in B(\alpha_\ell)$ the condition (7) is met

then this allotment is the classification result $R_c^*(X)$ for the value α_ℓ and stop

else construct the set of allotments $B'(\alpha_\ell) \subseteq B(\alpha_\ell)$ which satisfy condition (7) and go to step 8;

9. Perform the following operations for each allotment $R_{c(z)}^{\alpha_\ell}(X) \in B'(\alpha_\ell)$:

   9.1   Let $l := 1$;

   9.2   Find the support $Supp(A_{(\alpha_\ell)}^l) = A_{\alpha_\ell}^l$ of the fuzzy cluster $A_{(\alpha_\ell)}^l \in R_{c(z)}^{\alpha_\ell}(X)$ and construct the matrix of attributes $X_{n_l \times m} = [x_i^t]$, $x_i \in A_{\alpha_\ell}^l$, $t = 1,\ldots,m$ for $A_{\alpha_\ell}^l$ where $n_l = card(A_{\alpha_\ell}^l)$;

   9.3   Calculate the prototype $\tau^l = \{\bar{x}^1,\ldots,\bar{x}^m\}$ of the class $A_{\alpha_\ell}^l$ according to the formula $\bar{x}^t = \frac{1}{n_l}\sum_{x_i \in A_{\alpha_\ell}^l} x_i^t$, $t = 1,\ldots,m$;

   9.4   Calculate the distance $d(\tau^l, \tau^l)$ between the typical point τ^l of the fuzzy cluster $A_{(\alpha_\ell)}^l$ and its prototype τ^l;

   9.5   Check the following condition:

   if all fuzzy clusters $A_{(\alpha_\ell)}^l \in R_{c(z)}^{\alpha_\ell}(X)$ are not verified

   then let $l = l + 1$ and go to step 8.2
   else go to step 9

10. Compare the fuzzy clusters $A_{(\alpha_\ell)}^l$ which are elements of different allotments $R_{c(z)}^{\alpha_\ell}(X) \in B'(\alpha_\ell)$; the allotment $R_{c(z)}^{\alpha_\ell}(X) \in B'(\alpha_\ell)$ for which the distance $d(\tau^l, \tau^l)$ is minimal for all fuzzy clusters $A_{(\alpha_\ell)}^l$ is the classification result $R_c^*(X)$.

Notice that the minimal value of the tolerance threshold α^* is the parameter of the D-AFC-TAGA(α)-algorithm and this value should be estimated by an expert.

So, the D-AFC-TC-algorithm, the D-PAFC-TC-algorithm, and the D-AFC-TC(α)-algorithm are particular versions of the corresponding algorithms based on computing a transitive approximation $\hat{T}_k$, $k \in \{1,\ldots,6\}$ of the initial fuzzy tolerance T.

3 Experimental Results

An application of the proposed algorithms to the data processing should be explained by simple example. For the purpose, Sneath and Sokal's two-dimensional data set [6] was selected. The artificial data set is shown in Figure 1.

Let us consider an application of the D-AFC-TAGA-algorithm to Sneath and Sokal's data set for a case of manual choosing an aggregation operator $f_k(y_1, y_2,\ldots,y_m)$, $k \in \{1,\ldots,6\}$. The formula (8) and the squared normalized Euclidean distance (10) were used in the numerical experiment.
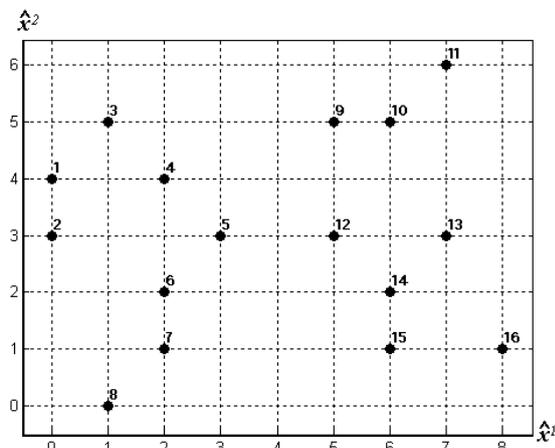
Fig. 1. Sneath and Sokal's data set

By executing the D-AFC-TAGA-algorithm for a case of using the maximum operator, we obtain the following: the first class is formed by 13 elements and the second class is composed of 3 elements. The allotment $R_c^*(X)$ among $c = 2$ fully separated fuzzy clusters, which corresponds to the result, is received for the tolerance threshold $\alpha = 0.96875$. 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. So, 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. Thus, the tenth object is the typical point of the second fuzzy cluster.
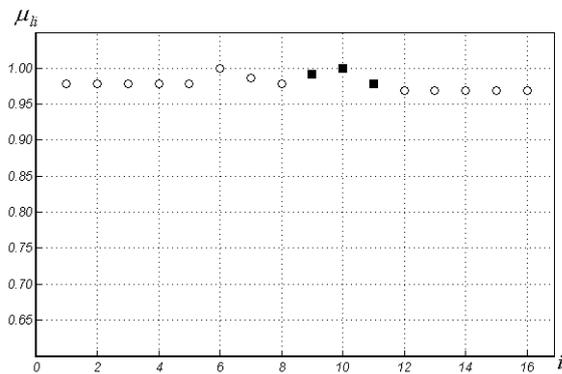
Fig. 2. Membership functions of two fuzzy clusters obtained from
the D-AFC-TAGA-algorithm by using the maximum operator
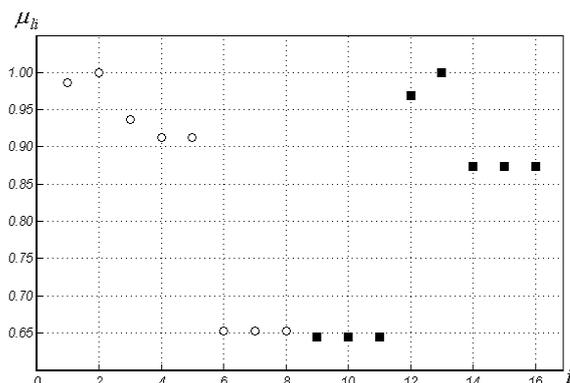
Fig. 3. Membership functions of two fuzzy clusters obtained from
the D-AFC-TAGA-algorithm by using the minimum operator

Membership functions of two classes of the allotment are presented in Figure 2 and values which equal zero are not shown in the figure. Membership values of the first class are represented by ○ and membership values of the second class are represented by ■.

The result obtained from the D-AFC-TAGA-algorithm by using the maximum operator is equal to the result obtained from the D-AFC-TC-algorithm [2], because the approximation which is generated by the maximum operator is the transitive closure of the initial fuzzy tolerance.

By executing the D-AFC-TAGA-algorithm for the minimum operator, we obtain the allotment $R_c^*(X)$ among $c = 2$ fully separated fuzzy clusters, which corresponds to the result, is received for the tolerance threshold $\alpha = 0.644965$. The first class is formed by 8 elements and the second class is also composed of 8 elements. The second object is the typical point of the first fuzzy cluster and the thirteenth object is the typical point of the second fuzzy cluster. Membership functions of two classes of the allotment are presented in Figure 3.

Results obtained from the D-AFC-TAGA-algorithm by using the mean operator, the median operator, the upmedian operator, and the downmedian operator are shown in Figures 4 – 7.
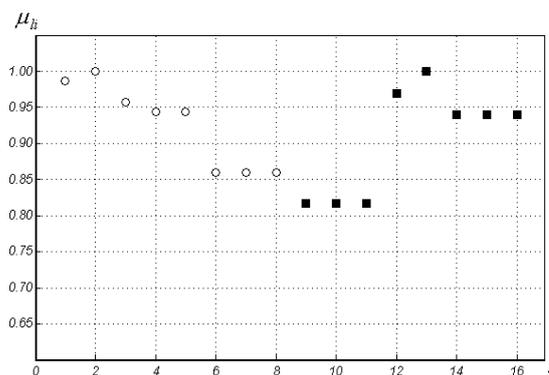
Fig. 4. Membership functions of two fuzzy clusters obtained from
the D-AFC-TAGA-algorithm by using the mean operator
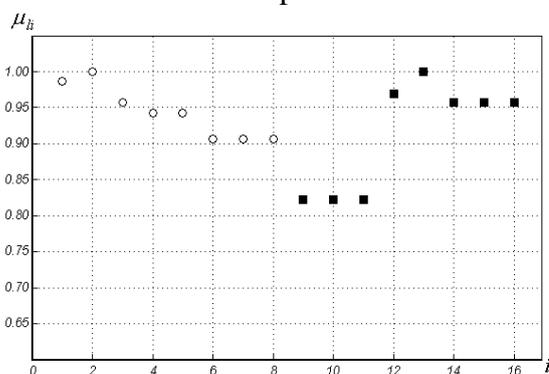
Fig. 5. Membership functions of two fuzzy clusters obtained from
the D-AFC-TAGA-algorithm by using the median operator
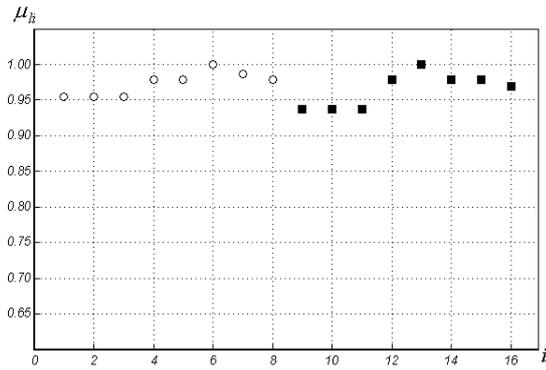
Fig. 6. Membership functions of two fuzzy clusters obtained from
the D-AFC-TAGA-algorithm by using the upmedian operator
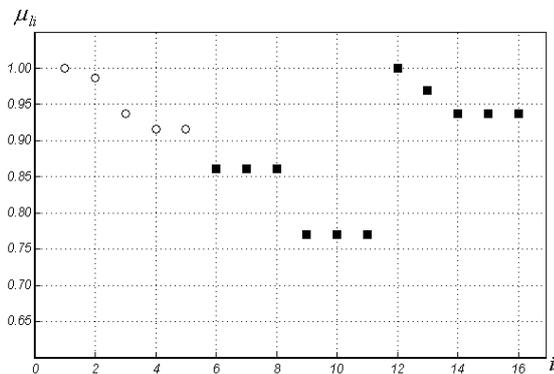
Fig. 7. Membership functions of two fuzzy clusters obtained from
the D-AFC-TAGA-algorithm by using the downmedian operator

Notice that in a case of using the downmedian operator in the TAGA-algorithm, the first class of the obtained allotment is formed by 5 elements and the second class is composed of 11 elements.

In a case of automatic constructing an acceptable transitive approximation $\tilde{T}_k$, $k = 1, \dots 6$ of the fuzzy tolerance T, the result obtained from the D-AFC-TAGA-algorithm is equal to the result obtained by using the mean operator. The fact is explained by Table 1 where values of the distance (11) are given. So, the condition $\min_k d(T, \tilde{T}_k)$, $k = 1, \dots 6$ is met for the transitive approximation obtained by using the mean operator. Moreover, Table 1 present also values of the tolerance threshold in obtained allotments.

Table 1. Values of the distance between fuzzy tolerance and its transitive approximations and values of the tolerance threshold for allotments obtained from different aggregation operators

A type of the aggregation operator	Values of the distance between fuzzy relations	Values of the tolerance threshold for the obtained allotment
maximum	39.43923	0.96875
minimum	76.63888	0.64496
mean	21.03686	0.81689
median	23.42621	0.82248
upmedian	29.02951	0.93663
downmedian	31.55729	0.76996

So, the results of classification obtained from the D-AFC-TAGA-algorithm seem to be appropriate than the results which were obtained from the D-AFC-TC-algorithm.

4 Conclusions
The family of direct prototype-based heuristic algorithms of possibilistic clustering for detection of unknown number of fuzzy clusters is proposed in the paper. The proposed algorithms based on constructing a transitive approximation of a fuzzy tolerance and these algorithms can be applied directly to the object by attributes data, by choosing a suitable distance between fuzzy sets to measure similarity. The result of application of the basic version of proposed algorithms to Sneath and Sokal's data seems to be satisfactory in comparison with the result obtained from the D-AFC-TC-algorithm. So, the proposed algorithms are more precise and effective tools for exploratory data analysis than the algorithms based on a transitive closure of a fuzzy tolerance.

References
[1] R. Krishnapuram and J.M. Keller, "A possibilistic approach to clustering", *IEEE Transactions on Fuzzy Systems*, vol. 1, no. 2, pp. 98-110, May 1993.
[2] D. A. Viattchenin, "A Heuristic Approach to Possibilistic Clustering: Algorithms

and Applications". Heidelberg: Springer, 2013.

[3] A. Kaufmann, "Introduction to the Theory of Fuzzy Subsets". New York: Academic Press, 1975.

[4] L. Garmendia and A. Salvador, "Computing a transitive opening of a reflexive and symmetric fuzzy relation", in *Proc. The Eighth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty ECS-QARU'2005*, Barcelona, Spain, 2005, pp. 587-599.

[5] P. Dawyndt, H. De Meyer, and B. De Baets, "UPGMA clustering revisited: A weight-driven approach to transitive approximation", *International Journal of Approximate reasoning*, vol. 42, no. 3, pp. 174-191, August 2006.

[6] P.H.A. Sneath and R. Sokal, "Numerical Taxonomy". San Francisco: Freeman, 1973.

Dmitri A. VIATTCHENIN was born on 13.12.1972 in Moscow, Russia. In 1994 he graduated the Department of Mathematical Modeling and Data Analysis of the Faculty of Applied Mathematics and Informatics in Belarusian State University. He defended his PhD in the field of philosophy of sciences and technology in 1998. He is currently a leading researcher in the Laboratory of Information Protection Problems of the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. His research interests include techniques of fuzzy and possibilistic clustering, fuzzy control, intelligent decision making systems, simulation of complex systems, and philosophical aspects of artificial intelligence. He is the author of over 130 publications including papers in international and domestic journals and conference proceedings, 3 monographs and 1 edited volume.

Aliaksandr DAMARATSKI was born on 29.11.1983 in Mogilev, Belarus. He has graduated the Faculty of Computer Design of Belarusian State University of Informatics and Radio-Electronics in 2005. He is currently a junior researcher in the Laboratory of Information Protection Problems of the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. His work focuses on possibilistic clustering, image processing, decision making, developing software, and computer networks analysis. He is the author of more than 20 papers in journals and conference proceedings.