A Heuristic Possibilistic Approach to Clustering for Asymmetric Data

Dmitri A. VIATTCHENIN
National Academy of Sciences of Belarus
United Institute of Informatics Problems
Surganov St. 6, 220012 Minsk, Belarus
viattchenin@mail.ru

Abstract—This paper deals with the problem of clustering of asymmetric data. A method of the problem solving is based on the application of a direct possibilistic clustering algorithm based on the concept of allotment among fuzzy cluster to a matrix of fuzzy tolerance, which correspond to the set of objects, for which asymmetric distances or proximities hold. The paper provides the description of the method of asymmetric data preprocessing for construction of a matrix of fuzzy tolerance and basic ideas of the method of clustering. An illustrative example of asymmetric data preprocessing and clustering is given and an analysis of the experimental results of the method’s application to the Sato-Ilic and Jain’s asymmetric data is carried out. Preliminary conclusions are discussed.

Keywords: Asymmetric data, possibilistic clustering, fuzzy preference relation, fuzzy tolerance.

I. INTRODUCTION

The aim of data mining is to discover knowledge from a large quantity of data. Cluster analysis is one of the well-known approaches in data mining. That is why clustering techniques have been actively studied.

Clustering is the process of assigning data objects into a set of disjoint groups called clusters so that objects in each cluster are more similar to each other than objects from different clusters. Algorithms of clustering can be broadly classified as hard, probabilistic, fuzzy, and possibilistic. From another hand, clustering methods can be divided into three types: heuristic methods, hierarchical methods and optimization methods.

Fuzzy and possibilistic clustering procedures have been applied effectively in image processing, data analysis, symbol recognition and modeling. 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 \( \mu_{li} \), \( 1 \leq l \leq c, 1 \leq i \leq n \), to which an object \( x_i \) belongs to the fuzzy cluster \( A' \) should be determined. For each object \( x_i, i=1,...,n \) the grades of membership should satisfy the conditions of a fuzzy \( c \)-partition:

\[
\sum_{l=1}^{c} \mu_{li} = 1, \ 1 \leq i \leq n; \ 0 \leq \mu_{li} \leq 1, \ 1 \leq l \leq c. \quad (1)
\]

So, the family of fuzzy sets \( P(X) = \{A' | l=1,c, c \leq n\} \) 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 \( P_{con} = [\mu_{li}], \ l=1,...,c, \ i=1,...,n \). The best known optimization approach to fuzzy clustering is the method of fuzzy \( c \)-means, developed by Bezdek [1].

However, the conditions (1) of fuzzy \( c \)-partition are very difficult from essential positions. So, if condition

\[
\sum_{i=1}^{c} \nu_{li} > 0, \ 1 \leq i \leq n; \ 0 \leq \nu_{li} \leq 1, \ 1 \leq l \leq c. \quad (2)
\]

is met for each object \( x_i, 1 \leq i \leq n \) then the corresponding family of fuzzy sets \( Y(X) = \{A' | l=1,c, c \leq n\} \) is the possibilistic partition of the set of objects \( X=\{x_1,...,x_n\} \). Membership values \( \nu_{li}, \ i=1,...,n, \ l=1,...,c \) can be interpreted as the values of typicality degree. The method of possibilistic \( c \)-means was proposed by Krishnapuram and Keller [2] and the \( PCM \)-algorithm method is most popular method in the possibilistic approach to clustering. Moreover, the possibilistic approach to clustering was developed by other researchers. A review of some optimization methods of fuzzy and possibilistic clustering was made by Höppner, Klawonn, Kruse and Runkler [3].

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 most popular examples of fuzzy relational clustering are the Windham’s \( AP \)-algorithm [4], the Hathaway, Davenport, and Bezdek’s \( RFDM \)-algorithm [5], and the \( ARCA \)-algorithm which was proposed by Corsini, Lazzzerini, and Marcelloni in [6].

Condition of symmetry of the initial data must be met in
the relational approach to fuzzy clustering. In other words, the object relation data matrix should satisfy the condition
\[ r(x_i, x_j) = r(x_j, x_i), \forall x_i, x_j \in X, \]  
where \( r(x_i, x_j) \) is either dissimilarity relation or similarity relation between the pair of different objects \( x_i \) and \( x_j \).

However, proximity data can be asymmetric in some cases. For example, a communication in human relationships, information flow and the degree of confusion based on perception for a discrimination problem are illustrative examples of asymmetric data. So, asymmetric data clustering methods are playing an important role in various areas of real applications. That is why clustering techniques based on asymmetric proximity have generated colossal interest among a number of researchers. In particular, some issues of clustering of the objects, for which asymmetric distances or proximities hold, are studied by Owsinski in [7].

From another hand, asymmetric clustering model, which was considered by Sato-Ilic and Jain in [8], is as follows:
\[ r_{ij} = \sum_{l=1}^{c} \sum_{m=1}^{c} w_{im}H_{jl}H_{mj} + e_{ij}, \]  
where \( r_{ij} = r(x_i, x_j), 0 \leq r_{ij} \leq 1 \) is the observed asymmetric proximity degree between objects \( x_i \) and \( x_j \). The grade \( r_{ij} \) does not equal \( r_{ji} \) when \( x_i \neq x_j \) and \( \mu_{ij} \) is a value which represents the grade of membership of an object \( x_i \) to a fuzzy cluster \( A_l \), \( l \in \{1, \ldots, c\} \) under the condition (1). In the model (4), the weight \( w_{im} \) is considered as a value which presents the asymmetric proximity between a pair of clusters \( A_l \) and \( A_m \). So, the assumption that the asymmetry of the proximity between the objects is caused by the asymmetry of the proximity between the clusters is met in the model (4).

However, the clustering technique based on the model (4) is complex from mathematical and essential positions. From other hand, if the observed data was obtained as \( n \times n \) asymmetric proximity matrix then the matrix can be interpreted as a preference relation on \( X \). The matrix of a fuzzy preference relation can be obtained after application of normalization to the matrix of initial asymmetric data and a fuzzy tolerance relation can be obtained from the fuzzy preference relation.

The goal of this paper is consideration of the heuristic possibilistic approach to clustering of asymmetric data based on the initial data transformation. The contents of this paper is as follows: in the second section definitions of fuzzy preference relation and fuzzy tolerance are given and a method of constructing of fuzzy tolerance from fuzzy preference relation is considered, in the third section basic concepts of the possibilistic clustering method based on the concept of allotment among fuzzy clusters are outlined, in the fourth section an illustrative example of the data preprocessing is shown and experimental results of application of the considered method to Sato-Ilic and Jain’s asymmetric data are given, in the fifth section some concluding remarks are stated and some perspectives are outlined.

II. FUZZY PREFERENCE RELATIONS AND FUZZY TOLERANCES

Let \( X = \{x_1, \ldots, x_n\} \) be a finite set of elements and \( R: X \times X \rightarrow [0,1] \) some binary fuzzy relation on \( X \) with \( \mu_R(x_i, x_j), \forall x_i, x_j \in X \) being its membership function. The weak fuzzy preference relation is the fuzzy binary relation which possesses the reflexivity property
\[ \mu_R(x_i, x_i) = 1, \forall x_i \in X. \]  
(5)

Condition
\[ \mu_R(x_i, x_j) \neq \mu_R(x_j, x_i), x_i \neq x_j, \forall x_i, x_j \in X \]  
(6)
is met for weak fuzzy preference relations very often.

Fuzzy tolerance \( T \) is the fuzzy binary intransitive relation on \( X \) which possesses the reflexivity property (5) and the symmetricity property
\[ \mu_T(x_i, x_j) = \mu_T(x_j, x_i), \forall x_i, x_j \in X. \]  
(7)

The notions of powerful fuzzy tolerance, feeble fuzzy tolerance and strict feeble fuzzy tolerance were considered in [9], as well. In this context the classical fuzzy tolerance in the sense of conditions (5) and (7) was called usual fuzzy tolerance. However, the essence of the heuristic possibilistic clustering method does not depend on the kind of fuzzy tolerance. The fuzzy tolerance \( T \), which corresponds to the weak fuzzy preference relation \( R \) on \( X \) can be constructed as follows [10]:
\[ \mu_T(x_i, x_j) = \mu_R(x_i, x_j) \land \mu_R(x_j, x_i), \forall x_i, x_j \in X. \]  
(8)

The transformation (8) is very important for the proposed technique of clustering of asymmetric data.

III. BASIC CONCEPTS OF THE HEURISTIC METHOD OF POSSIBILISTIC CLUSTERING

Heuristic algorithms of fuzzy clustering display high level of essential clarity and low level of a complexity. The algorithm of Chiang, Yue, and Yin [11] is a very good illustration for these characterizations. 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 [12] notes 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.

A heuristic approach to possibilistic clustering was outlined in [9] and developed in other publications. All
heuristic algorithms of possibilistic clustering are based on the concept of the allotment among fuzzy \(\alpha\)-clusters. Let us remind the basic concepts of the approach to clustering.

Let \(X = \{x_1, \ldots, x_n\}\) be the initial set of objects. Let \(T\) be a fuzzy tolerance on \(X\) and \(\alpha\) be \(\alpha\)-level value of \(T\), \(\alpha \in (0,1]\). Columns or lines of the fuzzy tolerance matrix are \(\alpha\)-level value of \(T\).

Let \(A'_\alpha = \{a'_\alpha\} = \{A'_\alpha(1), \ldots, A'_\alpha(n)\}\) be fuzzy sets on \(X\), which are generated by a fuzzy tolerance \(T\). The \(\alpha\)-level fuzzy set \(A'_\alpha = \{(x_i, \mu_\alpha(x_i))|\mu_\alpha(x_i) \geq \alpha, l \in [1,n]\}\) is fuzzy \(\alpha\)-cluster or, simply, fuzzy cluster. So \(A'_\alpha \subseteq A'_\alpha\), \(\alpha \in (0,1]\), \(A'_\alpha \in \{A'_\alpha, \ldots, A'_\alpha\}\) and \(u_\alpha\) is the membership grade of the element \(x_i \in X\) for some fuzzy cluster \(A'_\alpha\), \(\alpha \in (0,1]\), \(l \in [1,n]\). Value of \(\alpha\) is the tolerance threshold of fuzzy clusters elements.

The membership grade of the element \(x_i \in X\) for some fuzzy cluster \(A'_\alpha\), \(\alpha \in (0,1]\), \(l \in [1,n]\) can be defined as a

\[
u_\alpha = \begin{cases} \mu_\alpha(x_i), & x_i \in A'_\alpha \\ 0, & \text{otherwise} \end{cases},
\]

where an \(\alpha\)-level set \(A'_\alpha = \{x_i \in X|\mu_\alpha(x_i) \geq \alpha\}, \alpha \in (0,1]\) of a fuzzy set \(A'_\alpha\) is the support of the fuzzy cluster \(A'_\alpha\). So, condition \(A'_\alpha = \text{Supp}(A'_\alpha)\) is met for each fuzzy cluster \(A'_\alpha\), \(\alpha \in (0,1]\), \(l \in [1,n]\). Membership grade \(u_\alpha\) can be interpreted as a degree of typicality of an element \(x_i\) to a fuzzy cluster \(A'_\alpha\). The value of a membership function of each element of the fuzzy cluster in the sense of (9) is the grade of similarity of the object to some typical object of fuzzy cluster. Membership grade defines a possibility distribution function \(\pi_i(x_i)\) for some fuzzy cluster \(A'_\alpha\), \(\alpha \in (0,1]\).

Let \(T\) be a fuzzy tolerance on \(X\), where \(X\) is the set of elements, and \(\{A'_\alpha, \ldots, A'_\alpha\}\) is the family of fuzzy clusters for some \(\alpha \in (0,1]\). The point \(r'_\alpha \in A'_\alpha\), for which

\[
r'_\alpha = \arg \max_{x_i \in A'_\alpha} \nu_\alpha, \forall x_i \in A'_\alpha
\]

is called a typical point of the fuzzy cluster \(A'_\alpha\), \(\alpha \in (0,1]\), \(l \in [1,n]\). A fuzzy cluster can have several typical points.

Let \(R^\alpha(X) = \{A'_\alpha|\ l = \l_1, c, 2 \leq c \leq n\}\) be a family of fuzzy clusters for some value of tolerance threshold \(\alpha\), \(\alpha \in (0,1]\), which are generated by some fuzzy tolerance \(T\) on the initial set of elements \(X = \{x_1, \ldots, x_n\}\). If condition

\[
\sum_{i=1}^{n} \nu_\alpha > 0, \forall x_i \in X
\]

is met for all fuzzy clusters \(A'_\alpha \in R^\alpha(X)\), \(l = \l_1, c, c \leq n\), then the family is the allotment of elements of the set \(X = \{x_1, \ldots, x_n\}\) among fuzzy clusters \(\{A'_\alpha, l = \l_1, c, 2 \leq c \leq n\}\) for some value of the tolerance threshold \(\alpha\).

It should be noted that several allotments \(R^\alpha(X)\) can exist for some tolerance threshold \(\alpha\). That is why symbol \(z\) is the index of an allotment.

The condition (11) requires that every object \(x_i, i = 1, \ldots, n\) must be assigned to at least one fuzzy cluster \(A'_\alpha\), \(l = \l_1, c, c \leq n\) with the membership grade higher than zero. The condition \(2 \leq c \leq n\) requires that the number of fuzzy clusters in each allotment \(R^\alpha(X)\) must be more than two. Obviously, the definition of the allotment among fuzzy clusters (11) is similar to the definition of the possibilistic partition (2). So, the allotment among fuzzy clusters can be considered as the possibilistic partition and fuzzy clusters in the sense of (9) are elements of the possibilistic partition. However, the concept of allotment will be used in further considerations. The concept of allotment is the central point of the method. But the next concept introduced should be paid attention to, as well.

Allotment \(R^\alpha(X) = \{A'_\alpha|\ l = \l_1, n\}\) of the set of objects among \(n\) fuzzy clusters for some tolerance threshold \(\alpha \in (0,1]\) is the initial allotment of the set \(X = \{x_1, \ldots, x_n\}\).

If condition

\[
\sum_{i=1}^{n} \text{card}(A'_\alpha) \geq \text{card}(X), \forall A'_\alpha \in R^\alpha(X), \alpha \in (0,1],
\]

and condition

\[
\text{card}(A'_\alpha \cap A'_\alpha) \leq w, \forall A'_\alpha, A'_\alpha, \ l = m, \alpha \in (0,1]\]

are met for all fuzzy clusters \(A'_\alpha\), \(l = \l_1, c\) of some allotment \(R^\alpha(X) = \{A'_\alpha|\ l = \l_1, c, c \leq n\}\) then the allotment is the allotment among particularly separate fuzzy clusters and \(0 \leq w \leq n\) is the maximum number of elements in the intersection area of different fuzzy clusters. Obviously, if \(w = 0\) in conditions (12) and (13) then the intersection area of any pair of different fuzzy cluster is an empty set and fuzzy clusters are fully separate fuzzy clusters.

The adequate allotment \(R^\alpha(X)\) for some value of tolerance threshold \(\alpha \in (0,1]\) is a family of fuzzy clusters...
which are elements of the initial allotment $R^n_i(X)$ for the value of $\alpha$ and the family of fuzzy clusters should satisfy the conditions (12) and (13). The construction of adequate allotments $R^n_i(X) = \{A^i_{(\alpha)} | l = 1, c, c \leq n\}$ for every $\alpha \in (0,1]$ is a trivial problem of combinatorics.

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

$$F(R^n_i(X), \alpha) = \sum_{i=1}^{c} \frac{1}{n} \sum_{l=1}^{n} v_{n\beta} - \alpha \cdot c,$$

(14)

where $c$ is the number of fuzzy clusters in the allotment $R^n_i(X)$ and $n_l = \text{card}(A^i_{(\alpha)})$, $A^i_{(\alpha)} \in R^n_i(X)$ is the number of elements in the support of the fuzzy cluster $A^i_{(\alpha)}$, can be used for evaluation of allotments. Maximum of criterion (14) corresponds to the best allotment of objects among $c$ fuzzy clusters. So, the classification problem can be characterized formally as determination of the solution $R^*(X)$ satisfying

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

(15)

where $B = \{R^n_i(X)\}$ is the set of adequate allotments corresponding to the formulation of a classification problem.

Direct heuristic algorithms of possibilistic clustering can be divided into two types: prototype-based versus relational. The family of prototype-based clustering algorithms includes [13]

- D-AFC-TC-algorithm: constructing of the allotment among unknown number $c$ of fully separate fuzzy clusters;
- D-PAPC-TC-algorithm: constructing of the allotment among unknown minimal number $c$ of fully separate fuzzy clusters;
- D-AFC-TC($\alpha$)-algorithm: constructing of the allotment among unknown number $c$ of fully separate fuzzy clusters with respect to minimal value $\alpha^*$ of tolerance threshold.

These heuristic possibilistic clustering algorithms are based on transitive closure of fuzzy tolerance and the algorithms can be applied only to the matrix of attributes.

The matrix of the normalized data $X_{\text{norm}} = [x']$, $i = 1, \ldots, n, \ t = 1, \ldots, m$ is the matrix of the initial data for these algorithms. The allotment $R^*(X) = \{A^i_{(\alpha)} | l = 1, c\}$ among the unknown number $c$ of fuzzy clusters, the corresponding value of tolerance threshold $\alpha$, and fuzzy clusters prototypes are the results of classification.

From other hand, the group of relational algorithms includes

- D-AFC($\alpha$)-algorithm: constructing of the allotment among given number $c$ of partially separate fuzzy clusters [9];
- D-AFC-PS(c)-algorithm: partially supervised constructing of the allotment among given number $c$ of partially separate fuzzy clusters [14];
- D-PAPC-algorithm: constructing of the allotment among unknown minimal number $c$ of fully separate fuzzy clusters [15].

So, a matrix of fuzzy tolerance $T_{\text{norm}} = [\mu_\tau(x, x')]$, $i, j = 1, \ldots, n$ is the matrix of initial data for these direct heuristic algorithms of possibilistic clustering. The allotment $R^*(X) = \{A^i_{(\alpha)} | l = 1, c\}$ among either the given number or the unknown number $c$ of fuzzy clusters and the corresponding value of tolerance threshold $\alpha$ are the results of classification. Obviously, that the relational algorithms are useful for asymmetric data processing. In particular, the D-PAPC-algorithm will be used in further considerations.

### IV. A NUMERICAL EXAMPLE

#### A. The Data

The Sato and Jain’s asymmetric data originally appear in [8]. This data shows the human relations between 16 children. The value of this data shows the degree of disliking between the children. The number of clusters is determined as five. The original data are presented in Table I.

The Results of Classification

Let us consider the classification result using model (4) which was presented by Sato-Ilic and Jain. A fuzzy c-partition is the result of application of the model (4) to the asymmetric data. Membership functions $\mu_\beta$ of five classes $A^i, \ l \in \{1, \ldots, 5\}$ of the fuzzy c-partition are presented in Fig. 1 and values which equal zero are not shown in the figure.

Membership values of the first class are represented by $\circ$, membership values of the second class are represented by $\Delta$, membership values of the third class are represented by $\bullet$, membership values of the fourth class are represented by $\nabla$, and membership values of the fifth class are represented by $\sqcap$. 

---

90
TABLE I DISSIMILARITY MATRIX FOR THE CHILDREN

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

The children who are in the same cluster are good friends with each other. Clusters $A^1$ and $A^2$ consist of boys only, and clusters $A^3, A^4, A^5$ are girls only. For comparison, the D-PAFC-algorithm was applied to the transformed data. Let us consider results of experiments.

In the first place, a method of the data preprocessing must be considered. The matrix of observed data was normalized as follows:

$$r(x_j, x_k) = \frac{r(\hat{x}_j, \hat{x}_k)}{\max_{i,j} r(\hat{x}_i, \hat{x}_j)}.$$

(16)

where $r(\hat{x}_j, \hat{x}_k), i, j = 1, ..., 16$ are disliking coefficients between the children.

The weak fuzzy preference relation $R$ on $X = \{x_1, ..., x_{16}\}$ was constructed as follows:

$$\mu R(x_j, x_k) = 1 - r(x_j, x_k).$$

(17)

So, the fuzzy tolerance $T$ on $X$ was obtained from the weak fuzzy preference relation using the formula (8) The matrix of fuzzy tolerance $T_{i=16} = [\mu R(x_j, x_k)], i, j = 1, ..., 16$ was processed by the D-PAFC-algorithm and the allotment $R'(X)$ among five fully separate fuzzy clusters was obtained. The allotment, which corresponds to the result, was obtained for the tolerance threshold $\alpha = 0.6666$. Membership functions $\nu_0$ of five classes of the allotment are presented in Fig. 2 and values which equal zero are not shown in the figure. The fifth object is the typical point of the first fuzzy cluster, the second object is the typical point of the fuzzy cluster which corresponds to the second class, the eighth object is the typical point of the fuzzy cluster which corresponds to the third class, the tenth object is the typical point of the fourth fuzzy cluster, and the twentieth object is the typical point of the fuzzy cluster which corresponds to the fifth class. So, the results, which are obtained from the D-PAFC-algorithm using the proposed method of the asymmetric data preprocessing, are similar to the results, which were obtained by Sato-Ilic and Jain [8] using the model (4).

Moreover, the membership function from the proposed method is sharper than the membership function from the Sato-Ilic and Jain's method.
V. CONCLUDING REMARKS

The results of application of the possibilistic clustering method based on the allotment concept can be very well interpreted.

Moreover, the heuristic possibilistic clustering results are stable.

The methodology of heuristic possibilistic clustering of asymmetric data is outlined in the paper. The methodology is based on the initial data transformation and construction the fuzzy tolerance matrix. The results of application of the proposed methodology of the asymmetric data preprocessing and its processing by the D-PAFC-algorithm to the Sato-Ilic and Jain’s asymmetric data show that the methodology and the D-PAFC-algorithm are a precise and effective technique for the asymmetric data clustering. It should be noted that the proposed method is more general and simple, than of the method of fuzzy clustering of Sato-Ilic and Jain [8], because the latter method is more complex, than the here proposed method. Some other methods can be used for the initial asymmetric data preprocessing. These perspectives for investigations are of great interest both from the theoretical point of view and from the practical one as well.

ACKNOWLEDGMENT

I am grateful to Prof. Janusz Kacprzyk, Prof. Jan W. Owsiński, Prof. Frank Klawonn and Prof. Mika Sato-Ilic for their interest in my investigations and support. I also thank Mr. Aliaksandr Damaryski for elaborating experimental software. I would like to thank the anonymous referees for their valuable comments.

REFERENCES


