

# CONSTRUCTING STABLE CLUSTERING STRUCTURE FOR UNCERTAIN DATA SET

Dmitri A. VIATTCHENIN

Laboratory of Information Protection, United Institute of Informatics Problems, National Academy of Sciences of Belarus,  
Surganov Str. 6, 220012 Minsk, The Republic of Belarus, e-mail: viattchenin@mail.ru

## ABSTRACT

The paper deals with the problem of constructing the stable clustering structure for the uncertain data set. The problem of explaining of stability of the clustering structure in automatic classification of objects for varying attributes values is formulated. The proposed method of the uncertain data clustering is based on heuristic algorithms of possibilistic clustering. Basic concepts of the heuristic approach to possibilistic clustering based on the concept of allotment among fuzzy clusters, a validity measure and techniques of the data preprocessing are considered. A method of constructing the set of values of most possible number of fuzzy clusters for the uncertain data is provided and a technique of constructing the stable clustering structure is proposed. An illustrative example of the proposed technique application to the oil data set is carried out. An analysis of the experimental results is given and preliminary conclusions are formulated.

**Keywords:** possibilistic clustering, uncertain data, clustering structure, allotment, validity measure

## 1. INTRODUCTION

Some preliminary remarks are considered in the first subsection. Types of clustering structures are defined in the second subsection.

### 1.1. Preliminary remarks

The need for mechanisms that help to treat ambiguous, fuzzy and vague knowledge explains the grown-up interest in fuzzy systems. In particular, fuzzy clustering methods have been applied effectively in image processing, data analysis and modeling. 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 and the optimization methods of fuzzy clustering are based on the concept of fuzzy  $c$ -partition which is expressed as follows:

$$\sum_{l=1}^c u_{li} = 1, \quad 0 \leq u_{li} \leq 1, \quad (1)$$

where  $c$  is the number of fuzzy clusters  $A^l$ ,  $l = 1, \dots, c$  in the fuzzy  $c$ -partition  $P(X) = \{A^l \mid l = \overline{1, c}, c \leq n\}$  and  $u_{li}$ ,  $i = 1, \dots, n$  is the membership degree. So, the fuzzy  $c$ -partition can be arrayed as a  $(c \times n)$  matrix  $P = [u_{li}]$ . Objective function-based fuzzy clustering algorithms can be divided into two types: object versus relational. The best known object approach to fuzzy clustering is the method of fuzzy  $c$ -means [1]. From other hand, the most popular examples of fuzzy relational clustering are the RFCM-algorithm [2], and the ARCA-algorithm [3].

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. 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. Many authors have proposed several

measures of cluster validity associated with fuzzy  $c$ -partitions. For example, the partition coefficient is described in [1] and compactness and separation index was defined in [4]. The compactness and separation index is most popular cluster validity criteria. Notable, that the index is appropriate for the ARCA-algorithm, because the ARCA-algorithm, though being a relational clustering algorithm, generates prototypes.

A possibilistic approach to clustering was proposed by Krishnapuram and Keller in [5] and developed by other researchers. This approach can be considered as a way in the optimization approach in fuzzy clustering because all methods of possibilistic clustering are objective function-based methods. A concept of possibilistic partition is a basis of possibilistic clustering methods and membership values  $\mu_{li}$ ,  $l = 1, \dots, c$ ,  $i = 1, \dots, n$  can be interpreted as the values of typicality degree. For each object  $x_i$ ,  $i = 1, \dots, n$  the grades of membership should satisfy the conditions of a possibilistic partition:

$$\sum_{l=1}^c \mu_{li} > 0, \quad 0 \leq \mu_{li} \leq 1. \quad (2)$$

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  $X = \{x_1, \dots, x_n\}$  if condition (2) is met.

Objective function-based fuzzy clustering algorithms are the most widespread methods in fuzzy clustering. However, heuristic algorithms of fuzzy clustering display low level of a complexity and high level of essential clarity. 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. These algorithms are called algorithms of direct classification or direct clustering algorithms.

An outline for a new heuristic method of fuzzy clustering was presented in [6], where a basic version of direct clustering algorithm was described and the version of the algorithm is called the D-AFC(c)-algorithm [7]. The D-AFC(c)-algorithm can be considered as a direct algorithm of possibilistic clustering. The fact was

demonstrated in [7]. The heuristic approach to possibilistic clustering was developed, for example, in [8].

## 1.2. Types of clustering structures

Most fuzzy clustering techniques are designed for handling crisp data with their class membership functions. However, the data can be uncertain. Different types of uncertainty can be characterizing the initial data which must be processed by clustering algorithms. For example, a brief review of uncertain data clustering methods is given in [9]. An interval uncertainty of the initial data is a basic type of uncertainty in clustering problems. The interval-valued data is a particular case of the three-way data and the type of uncertainty is the subject of the consideration. So, a problem of the three-way clustering arises.

The problem of clustering the three-way data can be formulated as follows [9]. Let  $X = \{x_1, \dots, x_n\}$  is set of objects, where objects are indexed  $i$ ,  $i = 1, \dots, n$ ; each object  $x_i$  is described by  $m_1$  attributes, indexed  $t^1$ ,  $t^1 = 1, \dots, m_1$ , so that an object  $x_i$  can be represented by vector  $x_i = (x_i^1, \dots, x_i^{t_1}, \dots, x_i^{m_1})$ ; every attribute  $\hat{x}_i^{t_1}$ ,  $t_1 = 1, \dots, m_1$  can be characterized by  $m_2$  values of 2-ary attributes, so that  $\hat{x}_i^{t_1} = (\hat{x}_i^{t_1(1)}, \dots, \hat{x}_i^{t_1(t_2)}, \dots, \hat{x}_i^{t_1(m_2)})$ . So, the three-way data can be presented by a poly-matrix as follows:

$$\hat{X}_{n \times m_1 \times m_2} = [\hat{x}_i^{t_1(t_2)}], i = 1, \dots, n; \quad (3)$$

$$t_1 = 1, \dots, m_1, t_2 = 1, \dots, m_2.$$

In other words, the three-way data are the data, which are observed by the values of  $m_1$  attributes with respect to  $n$  objects for  $m_2$  situations. The purpose of the clustering is to classify the set  $X = \{x_1, \dots, x_n\}$  into  $c$  fuzzy clusters and the number of clusters  $c$  can be unknown because it is depend on the situation.

In the situation of interval uncertainty, the only information that we have about the actual value  $\hat{x}_i^{t_1}$  of some attribute  $\hat{x}_i^{t_1}$ ,  $t_1 \in \{1, \dots, m_1\}$  for the object  $x_i$ ,  $i \in \{1, \dots, n\}$  is that the value belongs to some interval, and if  $t_2 \in \{\min, \max\}$  then  $\hat{x}_i^{t_1} \in [\hat{x}_i^{t_1(\min)}, \hat{x}_i^{t_1(\max)}]$ . So,  $m_2 = 2$ , and the situation can be described by the expression  $\hat{x}_i^{t_1} = (\hat{x}_i^{t_1(\min)}, \hat{x}_i^{t_1(\max)})$ . Obviously, that if  $\hat{x}_i^{t_1(1)} = \dots = \hat{x}_i^{t_1(t_2)} = \dots = \hat{x}_i^{t_1(m_2)}$  for all  $t_1 = 1, \dots, m_1$ ,  $\forall i = 1, \dots, n$  then the initial data are the ordinary object data and it can be presented as the usual matrix of attributes,  $\hat{X}_{n \times m_1} = [\hat{x}_i^{t_1}]$ .

The initial data matrix (3) can be represented as a set of  $m_2$  matrices  $\hat{X}_{n \times m_1}^{t_2} = [\hat{x}_i^{t_1}]$ ,  $i = 1, \dots, n$ ,  $t_1 = 1, \dots, m_1$  and the "plausible" number  $c$  of fuzzy clusters can be different for each matrix  $\hat{X}_{n \times m_1}^{t_2} = [\hat{x}_i^{t_1}]$ ,  $t_2 \in \{1, \dots, m_2\}$ . A clustering structure of the data set depends on the type of the initial data.

Three types of the clustering structure were defined in [10]. Firstly, if the number of clusters  $c$  is some constant for each matrix  $\hat{X}_{n \times m_1}^{t_2} = [\hat{x}_i^{t_1}]$ ,  $t_2 \in \{1, \dots, m_2\}$  and the coordinates of prototypes  $\{\bar{\tau}^1, \dots, \bar{\tau}^c\}$  of the clusters  $\{A^1, \dots, A^c\}$  are constant, then the clustering structure called stable. Secondly, if the actual number of clusters  $c$  is some constant for each matrix  $\hat{X}_{n \times m_1}^{t_2} = [\hat{x}_i^{t_1}]$ ,  $t_2 \in \{1, \dots, m_2\}$  and the coordinates of prototypes of the clusters are not constant, then the clustering structure called quasi-stable. Thirdly, if the number of clusters  $c$  is different for matrices  $\hat{X}_{n \times m_1}^{t_2} = [\hat{x}_i^{t_1}]$ ,  $t_2 = 1, \dots, m_2$  then the clustering structure called unstable.

Detection of most "plausible" fuzzy clusters in the sought clustering structure for the uncertain data set  $X$  can be considered as a final aim of classification and the construction of the set of values of most possible number of fuzzy clusters with corresponding possibility degrees is the important step in this way.

The aim of the work is a detailed consideration of the method of the discovering the unique clustering structure, which corresponds to most natural allocation of objects among fuzzy clusters for the uncertain data set. In particular, the allocation among the most "plausible" unknown number of fuzzy clusters  $c$  must be detected.

So, the contents of this paper is as follows: in the second section basic concepts of the heuristic method of possibilistic clustering, a validity measure and techniques of the data preprocessing are considered, in the third section the method of constructing the stable clustering structure is proposed, in the fourth section an example of application of the proposed method to the Ichino and Yaguchi's oil data set [11] is given, in the fifth section some final remarks are stated.

## 2. HEURISTIC POSSIBILISTIC CLUSTERING

Basic concepts of the heuristic method of possibilistic clustering based on the allotment among fuzzy clusters concept are considered in the first subsection. A validity measure for the D-AFC(c)-algorithm is presented in the second subsection and techniques of the data preprocessing are given in the third subsection of the section.

### 2.1. Basic definitions

Direct heuristic algorithms of possibilistic clustering can be divided into two types: prototype-based [8] versus relational [6], [7]. The concept of fuzzy tolerance is the basis for the concept of fuzzy  $\alpha$ -cluster. That is why definition of fuzzy tolerance must be considered in the first place.

Let  $X = \{x_1, \dots, x_n\}$  be the initial set of elements and  $T: X \times X \rightarrow [0,1]$  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. Fuzzy tolerance is the fuzzy binary intransitive relation which possesses the symmetricity property

$$\mu_T(x_i, x_j) = \mu_T(x_j, x_i), \quad \forall x_i, x_j \in X, \quad (4)$$

and the reflexivity property

$$\mu_T(x_i, x_i) = 1, \forall x_i \in X. \tag{5}$$

Different fuzzy tolerances were considered in [6]. However, the essence of the method here considered does not depend on the kind of fuzzy tolerance and basic concepts are described for any fuzzy tolerance  $T$ .

Fuzzy similarity relation  $S$  is the fuzzy binary relation which possesses the symmetricity property (4), the reflexivity property (5), and the (max-min)-transitivity property:

$$\begin{aligned} \mu_S(x_i, x_k) &\geq \bigvee_{x_j \in X} (\mu_S(x_i, x_j) \wedge \mu_S(x_j, x_k)), \\ \forall x_i, x_j, x_k &\in X. \end{aligned} \tag{6}$$

Let some fuzzy binary relation be represented by a matrix  $R$  of size  $n$  and define

$$R^1 = R, R^n = R^{n-1} \circ R, n = 2, 3, \dots \tag{7}$$

The fuzzy binary relation  $R$  is a transitive closure of the relation  $R$  and  $\bar{R}$  is defined by the expression [12]

$$R = R^1 \cup R^2 \cup \dots \cup R^n, \tag{8}$$

where an operation  $\cup$  for two different fuzzy relations  $R^d$  and  $R^g$  is defined by the expression

$$\mu_{R^d \cup R^g}(x_i, x_j) = \mu_{R^d}(x_i, x_j) \vee \mu_{R^g}(x_i, x_j), \forall x_i, x_j \in X, \tag{9}$$

and the composition  $R^d \circ R^g$  of fuzzy relations  $R^d$  and  $R^g$  is defined in [12] as follows:

$$\begin{aligned} \mu_{R^d \circ R^g}(x_i, x_k) &= \bigvee_{x_j \in X} (\mu_{R^d}(x_i, x_j) \wedge \mu_{R^g}(x_j, x_k)), \\ \forall x_i, x_k &\in X. \end{aligned} \tag{10}$$

The transitive closure  $\bar{T}$  of some usual fuzzy tolerance  $T$  is a fuzzy similarity relation  $S$ . Let  $R$  denote a binary fuzzy relation. A level fuzzy relation  $R_{(\alpha)}$  is defined as

$$R_{(\alpha)} = \{(x_i, x_j) \in R, \mu_{R_{(\alpha)}}(x_i, x_j) = \mu_R(x_i, x_j)\} \mid \alpha \in (0, 1], \tag{11}$$

where  $R_\alpha$  is  $\alpha$ -level of the fuzzy relation  $R$  and  $\mu_R(x_i, x_j)$  is the membership function of the fuzzy relation  $R$ . The membership function of the level fuzzy relation  $R_{(\alpha)}$  for some  $\alpha$  can be defined as follows:

$$\mu_{R_{(\alpha)}}(x_i, x_j) = \begin{cases} \mu_R(x_i, x_j), & \text{if } \mu_R(x_i, x_j) \geq \alpha \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Let  $X = \{x_1, \dots, x_n\}$  be the initial set of objects. Let  $T$  be a fuzzy tolerance on  $X$  with  $\mu_T(x_i, x_j) \in [0, 1]$ ,  $\forall x_i, x_j \in X$  being its membership function and  $\alpha$  be the

$\alpha$ -level value of  $T$ ,  $\alpha \in (0, 1]$ . Columns or lines of the fuzzy tolerance matrix are fuzzy sets  $\{A^1, \dots, A^n\}$ . Let  $\{A^1, \dots, A^n\}$  be fuzzy sets on  $X$ , which are generated by a fuzzy tolerance  $T$ . The  $\alpha$ -level fuzzy set  $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha, x_i \in X\}$  is fuzzy  $\alpha$ -cluster or, simply, fuzzy cluster. So,  $A^l_{(\alpha)} \subseteq A^l$ ,  $\alpha \in (0, 1]$ ,  $A^l \in \{A^1, \dots, A^n\}$  and  $\mu_{A^l}$  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  $\alpha$  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

$$\mu_{A^l} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_{(\alpha)} \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where an  $\alpha$ -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  $A^l_{(\alpha)}$ .

The value of a membership function of each element of the fuzzy cluster is the degree of similarity of the object to some typical object of fuzzy cluster. Moreover, membership degree defines a possibility distribution function for some fuzzy cluster  $A^l_{(\alpha)}$ ,  $\alpha \in (0, 1]$ , and the possibility distribution function is denoted by  $\pi_l(x_i)$ . Notable that the number  $c$  of fuzzy clusters can be equal the number of objects,  $n$ .

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

$$\tau_e^l = \arg \max_{x_i} \mu_{A^l}, \forall x_i \in A^l_{(\alpha)}, \alpha \in (0, 1], \tag{14}$$

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

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

$$\sum_{l=1}^c \mu_{A^l} > 0, \forall x_i \in X \tag{15}$$

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

The condition (15) requires that every object  $x_i$ ,  $i = 1, \dots, n$  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. Obviously, the definition of the allotment among fuzzy clusters (15) is similar to the definition of the possibilistic partition (2).

Allotment  $R_l^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1, n}, \alpha \in (0, 1]\}$  of the set of objects among  $n$  fuzzy clusters for some threshold  $\alpha$  is the initial allotment of the set  $X = \{x_1, \dots, x_n\}$ . In other words, if a matrix of some level fuzzy tolerance in the sense of formulae (11) and (12) is given then lines or columns of the matrix are level fuzzy sets  $A_{(\alpha)}^l$ ,  $l = \overline{1, n}$ ,  $\alpha \in (0, 1]$  and the level fuzzy sets are fuzzy clusters. Membership functions of these fuzzy clusters are defined by the formula (13) for some value  $\alpha \in (0, 1]$ . So, these fuzzy clusters constitute an initial allotment for the tolerance threshold  $\alpha \in (0, 1]$  and they can be considered as clustering components.

If some allotment  $R_z^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1, c}, c \leq n\}$  corresponds to the formulation of a concrete problem, then this allotment is an adequate allotment. If condition

$$\sum_{l=1}^c \text{card}(A_{(\alpha)}^l) \geq \text{card}(X), \forall A_{(\alpha)}^l \in R_z^\alpha(X), \quad (16)$$

$$\alpha \in (0, 1], \text{card}(R_z^\alpha(X)) = c,$$

and condition

$$\text{card}(A_{(\alpha)}^l \cap A_{(\alpha)}^m) \leq w, \forall A_{(\alpha)}^l, A_{(\alpha)}^m, \quad (17)$$

$$l \neq m, \alpha \in (0, 1],$$

are met for all fuzzy clusters  $A_{(\alpha)}^l$ ,  $l = \overline{1, c}$  of some allotment  $R_z^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1, c}, c \leq n\}$  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 (16) and (17) then the intersection area of any pair of different fuzzy cluster is an empty set and fuzzy clusters are fully separate fuzzy clusters.

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

$$F(R_z^\alpha(X), \alpha) = \sum_{l=1}^c \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c, \quad (18)$$

where  $c$  is the number of fuzzy clusters in the allotment

$R_z^\alpha(X)$  and  $n_l = \text{card}(A_{(\alpha)}^l)$ ,  $A_{(\alpha)}^l \in R_z^\alpha(X)$  is the number of elements in the support of the fuzzy cluster  $A_{(\alpha)}^l$ , can be used for evaluation of allotments.

Maximum of criterion (18) 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

$$R^*(X) = \arg \max_{R_z^\alpha(X) \in B} F(R_z^\alpha(X), \alpha), \quad (19)$$

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

The condition (19) must be met for the some unique allotment  $R_z^\alpha(X) \in B$ . Otherwise, the number  $c$  of fuzzy clusters in the allotment sought  $R^*(X)$  is suboptimal.

Detection of fixed  $c$  number of particularly separate fuzzy clusters can be considered as the aim of classification. A general plan of the relational D-AFC(c)-algorithm is given, for example, in [6], [7] and [9]. The allotment  $R_z^\alpha(X) = \{A_{(\alpha)}^l \mid l = \overline{1, c}\}$  among the given number  $c$  of fuzzy clusters and the corresponding value of tolerance threshold  $\alpha \in (0, 1]$  are the results of classification obtained from the D-AFC(c)-algorithm.

From other hand, detection of the unique allotment  $R^*(X)$  among unknown number  $c$  of fully separated fuzzy clusters is the matter of the prototype-based D-AFC-TC-algorithm. The transitive closure  $T$  of some usual fuzzy tolerance  $T$  is constructed according to formulae (7) – (10). The fuzzy relation  $T$  is possesses properties (4), (5) and (6). The transitive closure is used in the clustering procedure and an idea of a leap in ordered sequence  $0 < \alpha_0 < \dots < \alpha_\ell < \dots < \alpha_z \leq 1$  of the tolerance threshold values for finding of the appropriate value  $\alpha_\ell$  are the basis of the algorithm. The matrix of attributes  $\hat{X}_{n \times m_1} = [\hat{x}_i^j]$  is the matrix of initial data and some distance  $d(x_i, x_j)$  for fuzzy sets [12] is a parameter for the D-AFC-TC-algorithm. A plan of the D-AFC-TC-algorithm is presented in [8]. The allotment  $R^*(X)$  among the unknown number  $c$  of fully separated fuzzy clusters, the corresponding value of tolerance threshold  $\alpha$  and normalized coordinates of prototypes  $\bar{\tau}^l$  of fuzzy clusters  $A_{(\alpha)}^l \in R^*(X)$ ,  $l = 1, \dots, c$  are results of classification.

## 2.2. A validity measure

Let  $R_c^*(X)$  be the allotment which is corresponds to the result of classification for the given number  $c$  of fuzzy clusters,  $c \in \{2, \dots, n\}$ . The quadratic measure of fuzziness of the allotment was defined in [13] as follows:

$$V_{OMF}(R_c^*(X); c) = \sum_{A_{(\alpha)}^l \in R_c^*(X)} \left( \frac{2}{\sqrt{n_l}} \cdot d_E(A_{(\alpha)}^l, \bar{A}_{(\alpha)}^l) \right), \quad (20)$$

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

$$d_E(A_{(\alpha)}^i, \underline{A}_{(\alpha)}^i) = \left( \sum_{x_i \in A_{(\alpha)}^i} (\mu_{A_{(\alpha)}^i}(x_i) - \mu_{\underline{A}_{(\alpha)}^i}(x_i))^2 \right)^{\frac{1}{2}}, \quad (21)$$

between the fuzzy cluster  $A_{(\alpha)}^i$  and the crisp set  $\underline{A}_{(\alpha)}^i$  nearest to the fuzzy cluster  $A_{(\alpha)}^i$ . The membership function of the crisp set  $\underline{A}_{(\alpha)}^i$  in the equation (21) can be defined as

$$\mu_{\underline{A}_{(\alpha)}^i}(x_i) = \begin{cases} 0, & \mu_{A_{(\alpha)}^i}(x_i) \leq 0.5, \\ 1, & \mu_{A_{(\alpha)}^i}(x_i) > 0.5, \end{cases} \quad \forall x_i \in A_{(\alpha)}^i, \quad (22)$$

where  $A_{(\alpha)}^i = \text{Supp}(A_{(\alpha)}^i)$  and  $\alpha \in (0,1]$ . Using the validity quadratic measure (20) the optimal number of fuzzy clusters can be obtained by maximizing the index value.

**2.3. Notes on the data preprocessing**

The D-AFC(c)-algorithm can be applied directly to the data given as a matrix of fuzzy tolerance  $T = [\mu_T(x_i, x_j)]$ ,  $i, j = 1, \dots, n$ . This means that it can be used with the objects by attributes data by choosing a suitable metric to measure similarity. The three-way data can be normalized as follows [9]:

$$x_i^{t_1(t_2)} = \frac{\hat{x}_i^{t_1(t_2)} - \min_{i,t_2} \hat{x}_i^{t_1(t_2)}}{\max_{i,t_2} \hat{x}_i^{t_1(t_2)} - \min_{i,t_2} \hat{x}_i^{t_1(t_2)}}. \quad (23)$$

So, each object  $x_i$ ,  $i = 1, \dots, n$  from the initial set  $X = \{x_1, \dots, x_n\}$  can be considered as a type-two fuzzy set and  $x_i^{t_1(t_2)} = \mu_{x_i}(x^{t_1(t_2)})$ ,  $i = 1, \dots, n$ ;  $t_1 = 1, \dots, m_1$ ,  $t_2 = 1, \dots, m_2$ ,  $x^{t_1(t_2)} = \mu_{x_i}(x^{t_2}) \in [0,1]$ ,  $t_1 = 1, \dots, m_1$ ,  $t_2 = 1, \dots, m_2$  are its membership functions. In the case of three-way data each object  $x_i$ ,  $i = 1, \dots, n$  can be presented as a matrix  $X_{(i)m_1 \times m_2} = [x_i^{t_1(t_2)}]$ ,  $t_1 = 1, \dots, m_1$ ,  $t_2 = 1, \dots, m_2$ .

Dissimilarity coefficients between the objects can be constructed on a basis of generalizations of distances between fuzzy sets [9] and these generalizations are taken into account dissimilarities between objects attributes as well as attributes situations. In particular, a generalization of the squared normalized Euclidean distance for type-two fuzzy sets is described by the expression

$$\varepsilon_G(x_i, x_j) = \frac{1}{m_1} \sum_{t_1=1}^{m_1} \left( \frac{1}{m_2} \sum_{u_i, v_i=1}^{m_2} (\mu_{x_i}(x^{t_1(u_i)}) - \mu_{x_j}(x^{t_1(v_i)}))^2 \right), \quad (24)$$

for all  $i, j = 1, \dots, n$ . In the case  $m_2 = 1$ , the formula (24) will be rewritten as the usual squared normalized

Euclidean distance [12]:

$$\varepsilon(x_i, x_j) = \frac{1}{m_1} \sum_{t_1=1}^{m_1} (\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1}))^2, \quad i, j = 1, \dots, n. \quad (25)$$

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

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j), \quad \forall i, j = 1, \dots, n, \quad (26)$$

to the matrix of fuzzy intolerance  $I = [\mu_I(x_i, x_j)]$ .

From other hand, in the case of the interval-valued data  $m_2 = 2$  and  $t_2 \in \{\min, \max\}$ . Thus, each object  $x_i$ ,  $i = 1, \dots, n$  can be considered as an interval-valued fuzzy set and  $\mu_{x_i}(x^{t_1}) = [\mu_{x_i}(x^{t_1(\min)}), \mu_{x_i}(x^{t_1(\max)})]$ ,  $i = 1, \dots, n$ ,  $t_1 = 1, \dots, m_1$  is its membership function, where  $\mu_{x_i}(x^{t_1(\min)}) \in [0,1]$ ,  $\mu_{x_i}(x^{t_1(\max)}) \in [0,1]$ . So, the formula (24) can be rewritten as follows:

$$\varepsilon_I(x_i, x_j) = \frac{1}{m_1} \sum_{t_1=1}^{m_1} \left( \frac{1}{2^2} \sum_{t_2 \in \{\min, \max\}} (\mu_{x_i}(x^{t_1(t_2)}) - \mu_{x_j}(x^{t_1(t_2)}))^2 \right), \quad (27)$$

for all  $i, j = 1, \dots, n$ .

Different distances and similarity measures for interval-valued fuzzy sets were proposed in other publications [14], [15]. For example, a similarity measure was defined by Ju and Yuan in [14] as follows:

$$s_I(x_i, x_j) = 1 - \frac{1}{\sqrt[\lambda]{m_1}} \sqrt[\lambda]{ \frac{\frac{\mu_{x_i}(x^{t_1(\min)}) + \mu_{x_i}(x^{t_1(\max)})}{2}}{\frac{\mu_{x_j}(x^{t_1(\min)}) + \mu_{x_j}(x^{t_1(\max)})}{2}} }, \quad (28)$$

for all  $i, j = 1, \dots, n$  and  $1 \leq \lambda < \infty$ .

From other hand, the normalized Euclidean distance between interval-valued fuzzy sets based on Hausdorff metric was defined by Grzegorzewski in [15] as follows:

$$e_I(x_i, x_j) = \sqrt{\frac{1}{m_1} \sum_{t_1=1}^{m_1} \max \left\{ \begin{aligned} & (\mu_{x_i}(x^{t_1(\min)}) - \mu_{x_j}(x^{t_1(\min)}))^2, \\ & (\mu_{x_i}(x^{t_1(\max)}) - \mu_{x_j}(x^{t_1(\max)}))^2 \end{aligned} \right\}}, \quad (29)$$

for all  $i, j = 1, \dots, n$ .

**3. THE PROPOSED METHOD**

A procedure for constructing the set of values of most possible number of fuzzy clusters in some sought structure is described in the first subsection. The second subsection includes the technique of constructing the stable clustering structure.

### 3.1. Constructing the set of values of the most possible number of fuzzy clusters

Let us remember the concept of fuzzy number which is useful for constructing the set of values of most possible number of fuzzy clusters [16].

Let  $L$  or  $R$  be decreasing, shape functions from  $\mathfrak{R}^+$  to  $[0,1]$  with  $L(0)=1$  and  $\forall u > 0, L(u) < 1, \forall u < 1, L(u) > 0; L(1)=0$  or  $L(u) > 0, \forall u$  and  $L(+\infty)=0$ . Then a fuzzy set  $V$  is called a LR-type fuzzy number  $V=(m,a,b)_{LR}$  with  $a > 0, b > 0$  if a membership function  $\mu_V(u)$  of  $V$  is defined as

$$\mu_V(u) = \begin{cases} L\left(\frac{m-u}{a}\right), & \text{for } u \leq m, \\ R\left(\frac{u-m}{b}\right), & \text{for } u \geq m, \end{cases} \quad (30)$$

where  $m$  is called the mean value of  $V$  and  $a$  and  $b$  are called the left and right spreads. In LR-type fuzzy numbers, the triangular and Gaussian fuzzy numbers are most commonly used.

The concept of S-norm is also important for the consideration. If  $A$  and  $B$  are fuzzy sets in a universe  $U$  and  $\mu_A(u), \mu_B(u)$  their membership functions, then the fuzzy union  $C$  of  $A$  and  $B$  has the membership function  $\mu_C(u) = S(\mu_A(u), \mu_B(u))$ . Thus, an S-norm is sometimes called a fuzzy union. Depending on the choice of an S-norm, we obtain different fuzzy unions. For example, the most widely used S-norms are [16]:

- Maximum operation:

$$S_1(\mu_A(u), \mu_B(u)) = \max(\mu_A(u), \mu_B(u)), \quad (31)$$

- Probabilistic sum:

$$S_2(\mu_A(u), \mu_B(u)) = \mu_A(u) + \mu_B(u) - \mu_A(u) \cdot \mu_B(u), \quad (32)$$

- Bounded sum:

$$S_3(\mu_A(u_g), \mu_B(u_g)) = \min(1, \mu_A(u_g) + \mu_B(u_g)). \quad (33)$$

There is a five-step procedure for constructing the set of values of most possible number of fuzzy clusters:

1. The set of  $m_2$  matrices  $X_{n \times m_1}^{t_2} = [x_i^{t_1}]$  can be constructed after normalizing the initial data matrix  $\hat{X}_{n \times m_1 \times m_2} = [\hat{x}_i^{t_1(t_2)}]$  and each matrix should be processed by the D-AFC-TC-algorithm choosing a suitable distance  $d(x_i, x_j)$ ; the number  $c^{(t_2)}$  of fully separated fuzzy clusters in the allotment  $R^*(X^{t_2})$  and the corresponding value of tolerance threshold  $\alpha^{(t_2)}$  are main results of classification and the results can be described by two sets  $\tilde{C} = \{c^{(t_2)} \mid \hat{t}_2 = 1, \dots, \hat{m}_2\}$  and  $\tilde{A} = \{\alpha^{(t_2)} \mid \hat{t}_2 = 1, \dots, \hat{m}_2\}$  where  $1 \leq \hat{m}_2 \leq m_2$ ;
2. Construct the LR-type fuzzy number  $V^{(t_2)} = (m^{(t_2)}, a^{(t_2)}, b^{(t_2)})_{LR}$  for every value  $c^{(t_2)} \in \tilde{C}$  as

follows:  $m^{(t_2)} = c^{(t_2)}, a^{(t_2)} = m^{(t_2)} - 1, b^{(t_2)} = n - m^{(t_2)}$  and membership functions of fuzzy numbers  $V^{(t_2)}, \hat{t}_2 \in \{1, \dots, \hat{m}_2\}$  are defined by expression (30) where  $u \in (1, n)$  and  $\mu_{V^{(t_2)}}(1) = \mu_{V^{(t_2)}}(n) = 0$ ;

3. Construct the fuzzy set  $\hat{V}^{(t_2)} = \{\hat{c}_g, \mu_{\hat{V}^{(t_2)}}(\hat{c}_g)\}$  from each LR-type fuzzy number  $V^{(t_2)}, \hat{t}_2 \in \{1, \dots, \hat{m}_2\}$  as follows: a subset of integer values  $\hat{C} = \{\hat{c}_*, \dots, \hat{c}^*\}$  where  $\hat{c}_* = 2$  and  $\hat{c}^* = n - 1$  should be extracted from the continuum  $(1, n)$  and the value of the membership degree  $\mu_{\hat{V}^{(t_2)}}(\hat{c}_g), \hat{c}_g \in \hat{C}$  of each fuzzy set  $\hat{V}^{(t_2)}$  is equal to the membership function value  $\mu_{V^{(t_2)}}(u)$  of corresponding fuzzy number  $V^{(t_2)}$  in the case  $u = \hat{c}_g$ ;
4. Construct the fuzzy union  $D = S_{\hat{t}_2}(\mu_{\hat{V}^{(t_2)}}(\hat{c}_g))$  of all fuzzy sets  $\hat{V}^{(t_2)} = \{\hat{c}_g, \mu_{\hat{V}^{(t_2)}}(\hat{c}_g)\}, \hat{c}_g \in \hat{C}, \hat{t}_2 = 1, \dots, \hat{m}_2$  with its membership function  $\mu_{\hat{D}}(\hat{c}_g)$  according to a selected S-norm;
5. Construct the  $\alpha$ -level fuzzy set for the fuzzy set  $D$  as follows:  $D_{(\hat{\alpha})} = \{c_g \in D_{\alpha}, \mu_{D_{(\hat{\alpha})}}(\hat{c}_g) = \mu_D(\hat{c}_g)\}$ , where  $D_{\hat{\alpha}} = \{\hat{c}_g \in \hat{C} \mid \mu_{\hat{D}}(\hat{c}_g) \geq \hat{\alpha}\}$  and  $\hat{\alpha} = \min_{\hat{t}_2} \alpha^{(t_2)}$ .

So, the set  $D_{\hat{\alpha}} = \text{Supp}(D_{(\hat{\alpha})})$  is the set of values of most possible number of fuzzy clusters in some sought clustering structure. The membership function  $\mu_{D_{(\hat{\alpha})}}(\hat{c}_g)$  can be interpreted as a possibility distribution  $\pi$  [16] and possibility degrees  $\pi(\hat{c}_g)$  express the extent to which the number  $c_g \in D_{\hat{\alpha}}$  of fuzzy clusters is plausible.

### 3.2. Constructing the stable clustering structure

A technique of constructing the stable clustering structure for the uncertain data set can be considered as a two-step process, where the set  $D_{\hat{\alpha}}$  of values of most possible number of fuzzy clusters is a preliminary result of classification. The allotment  $R^*(X)$  among a priori unknown number of fuzzy clusters can be considered as the sought clustering structure. So, the technique of constructing the allotment among fuzzy clusters for the uncertain data set can be summarized as follows:

1. The initial data are contained in the poly-matrix of attributes  $\hat{X}_{n \times m_1 \times m_2} = [\hat{x}_i^{t_1(t_2)}], i = 1, \dots, n, t_1 = 1, \dots, m_1, t_2 = 1, \dots, m_2$  and the procedure of constructing the set  $D_{\hat{\alpha}}$  of values of most possible number of fuzzy clusters should be applied to the data set;
2. The matrix of tolerance coefficients  $T = [\mu_T(x_i, x_j)], i, j = 1, \dots, n$  can be constructed from the normalized initial data by choosing a suitable distance for type-two or interval-valued fuzzy sets;
3. The D-AFC(c)-algorithm using some cluster validity index can be applied directly to the matrix of tolerance

coefficients for the set  $D_{\hat{\alpha}}$  and the allotment  $R^*(X)$  is the final result of the classification process.

The proposed technique can be generalized for a case of the fuzzy  $c$ -partition (1) very simply. An application of the proposed technique to classification problem will be illustrated on the interval-valued data example in the next section.

**4. AN ILLUSTRATIVE EXAMPLE**

The Ichino and Yaguchi’s interval-valued oil data set is described in the first subsection and results of the data processing are presented in the second subsection.

**4.1. The oil data set**

Let us consider the set of interval-valued data [11] which is presented in Table 1. The data set consists of the specific gravity, iodine value, and saponification value measured for 8 types of oils.

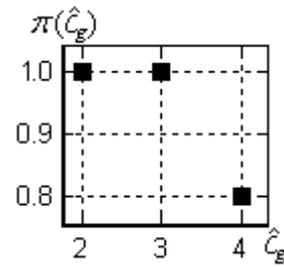
**Table 1** The Ichino and Yaguchi’s oil data set

Oils	Specific gravity	Iodine value	Saponification value
Linseed oil	0.930-0.935	170-204	118-196
Perilla oil	0.930-0.935	192-208	188-197
Cottonseed oil	0.916-0.918	99-113	189-198
Sesame oil	0.920-0.926	104-116	187-193
Camellia oil	0.916-0.917	80-82	189-193
Olive oil	0.914-0.919	79-90	187-196
Beef tallow	0.860-0.870	40-48	190-199
Hog fat	0.858-0.864	53-77	190-202

The analysis of types of oils in Table 1 highlights that the first six oils are vegetable and the remaining two are animal. That is why we expect to find two clusters in this data set.

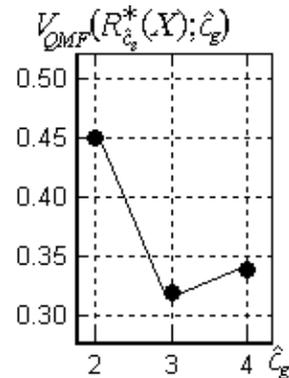
**4.2. Experimental results**

The data matrix  $\hat{X}_{8 \times 3 \times 2} = [\hat{x}_i^{(t_2)}]$ ,  $t_2 \in \{\min, \max\}$  was normalized according to formula (23). The procedure for constructing the set of values of most possible number of fuzzy clusters was applied to the normalized data using the squared normalized Euclidean distance (25) for the D-AFC-TC-algorithm and the maximum operation (31) as the fuzzy union. So, the set of values of most possible number of fuzzy clusters in the sought allotment  $R^*(X)$  is  $D_{\hat{\alpha}} = \{2, \dots, 4\}$  and corresponding possibility degrees are shown in Fig. 1.



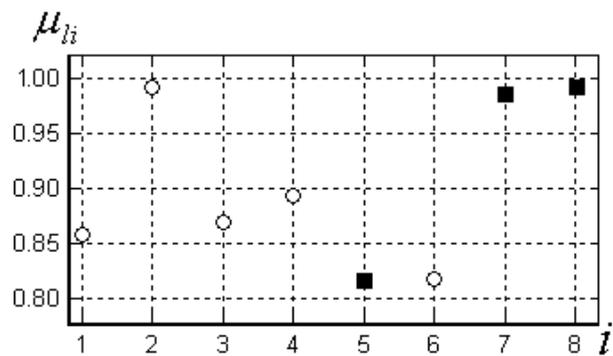
**Fig. 1** Possibility degrees constructed according to the maximum operation

The set  $D_{\hat{\alpha}}$  and possibility degrees  $\pi(\hat{c}_g)$  depend on the kind of fuzzy union. If the probabilistic sum (32) or the bounded sum (33) are selected as the fuzzy union then the set of values of most possible number of fuzzy clusters is  $D_{\hat{\alpha}} = \{2, \dots, 5\}$ . By executing the D-AFC(c)-algorithm for  $\hat{c}_g \in D_{\hat{\alpha}} = \{2, \dots, 4\}$  using the formula (27) in the process of the initial data preprocessing and the quadratic measure of fuzziness of the allotment (20) in the process of the classification performance, we obtain that the actual number of fuzzy clusters is equal 2 and the number corresponds to the maximum of the quadratic measure of fuzziness of the allotment (20). The performance of the validity measure is shown in Fig. 2.



**Fig. 2** Plot of the quadratic measure of fuzziness of the allotment as a function of the number of clusters using the formula (27)

Membership functions of two classes of the allotment  $R^*(X)$  are presented in Fig. 3.



**Fig. 3** Membership functions of two classes obtained from the D-AFC(c)-algorithm using the formula (27)

The allotment  $R^*(X)$  among two fully separated fuzzy clusters was obtained for the tolerance threshold  $\alpha = 0.8184$ . The first class is formed by 5 elements and the second class includes 3 elements. So, the fifth element is the misclassified object. Membership values of the first class are represented in Fig. 3 by  $\circ$ , membership values of the second class are represented by  $\blacksquare$  and values which equal zero are not shown in the figure.

By executing the D-AFC(c)-algorithm for  $\hat{c}_g \in D_{\hat{c}} = \{2, \dots, 4\}$  using the Ju and Yuan's similarity measure (28) for  $\lambda = 2$  and the validity measure (20), we obtain that numbers  $c = 3$  and  $c = 4$  of fuzzy clusters in the allotment sought  $R^*(X)$  are suboptimal. Thus, the allotment  $R^*(X)$  among two fully separated fuzzy clusters was obtained for the tolerance threshold  $\alpha = 0.5852$ . Membership functions of two classes of the allotment are presented in Fig. 4.

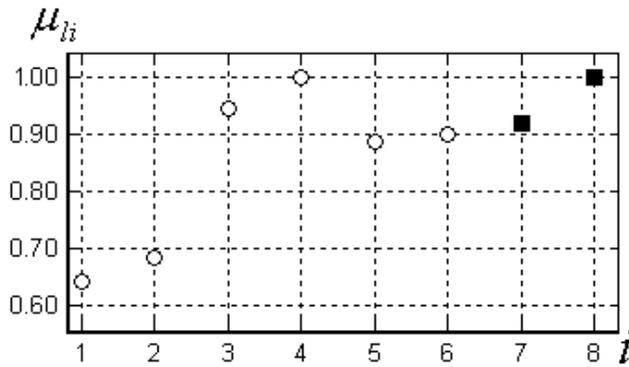


Fig. 4 Membership functions of two classes obtained from the D-AFC(c)-algorithm using the formula (28)

The first class is formed by 6 elements and the second class includes 2 elements. So, misclassified objects are absent at the resulting allotment  $R^*(X)$ .

By executing the D-AFC(c)-algorithm using the distance (29) and the validity measure (20), we obtain that the number  $c = 4$  of fuzzy clusters in the allotment sought  $R^*(X)$  is suboptimal and the actual number of fuzzy clusters is equal 2. The performance of the validity measure is shown in Fig. 5.

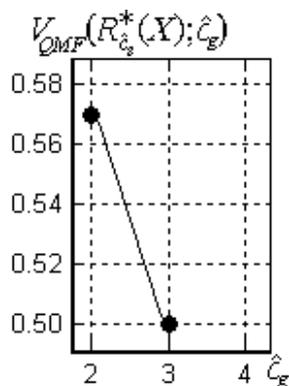


Fig. 5 Plot of the quadratic measure of fuzziness of the allotment as a function of the number of clusters using the formula (29)

Membership functions of two classes of the allotment are presented in Fig. 6, where membership values of the first class are represented by  $\circ$ , membership values of the second class are represented by  $\blacksquare$  and values which equal zero are not shown in the figure.

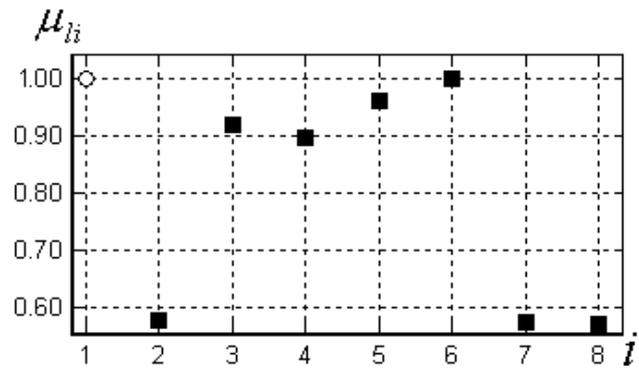


Fig. 6 Membership functions of two classes obtained from the D-AFC(c)-algorithm using the formula (29)

The first class is formed by 1 elements and the second class includes 7 elements. So, five misclassifications are presented in the resulting allotment  $R^*(X)$ .

In order to compare the presented results with the results obtained from the relational ARCA-algorithm of fuzzy clustering [3], we observed that the minimal value of the compactness and separation index [4] corresponds to the three fuzzy clusters and the maximal value of the partition coefficient [1] corresponds to the four fuzzy clusters in the fuzzy  $c$ -partition. So, presented results of the proposed technique of constructing the stable clustering structure seem to be appropriate.

5. FINAL REMARKS

Preliminary conclusions are discussed in the first subsection of the section. The second subsection deals with the perspectives on future investigations.

5.1. Conclusions

The technique of constructing the stable clustering structure for the uncertain data set is proposed in the paper. The results of application of the proposed technique to the oil data set show that the technique is the effective tool for solving the classification problem under uncertainty of the initial data.

Two heuristic algorithms of possibilistic clustering are the basis of the proposed technique. However, the results obtained from these algorithms depend on the selection of the dissimilarity measure and the initial data normalization method. Moreover, the set of values of most possible number of fuzzy clusters in the sought clustering structure depends on the type of the selected S-norm. So, we can conclude that the use of some one dissimilarity measure may produce serious hesitation. It will be a reasonable way to make use of various dissimilarity measures and compare the obtained clustering results.

The allocation of objects among the a priori unknown number of fuzzy clusters, which is the result of application of the proposed technique to the initial uncertain data set,

is appropriate for any actual values of the measured quantities  $\hat{x}_i^{t_2}$  belongs either to the interval  $[\hat{x}_i^{t_1(\min)}, \hat{x}_i^{t_1(\max)}]$ ,  $t_2 \in \{\min, \max\}$  in the situation of interval uncertainty, or to the set  $(\hat{x}_i^{t_1(1)}, \dots, \hat{x}_i^{t_1(t_2)}, \dots, \hat{x}_i^{t_1(m_2)})$ ,  $t_2 \in \{1, \dots, m_2\}$  in the case of the three-way data for all objects  $x_i \in X$ ,  $t_1 = 1, \dots, m_1$ . So, the obtained allotment  $R^*(X)$  among a discovered number  $c$  of fuzzy clusters can be considered as the stable clustering structure.

## 5.2. Perspectives

Firstly, possibility degrees for values of the most possible number of fuzzy clusters in some sought clustering structure can be taken into account in the classification process.

Secondly, a method for deriving fuzzy classification rules from the interval-valued data was outlined in [17]. The method is based on heuristic possibilistic clustering the interval-valued data. So, the technique proposed here can be considered as a preliminary step in the method of constructing a base of fuzzy rules.

## ACKNOWLEDGMENT

I am grateful to Prof. Janusz Kacprzyk, Prof. Jan W. Owsinski and Prof. Frank Klawonn for their interest in my investigations and support. I would like to thank the anonymous referees for their valuable comments.

## REFERENCES

- [1] BEZDEK, J. C.: *Pattern Recognition with Fuzzy Objective function Algorithms*. New York: Plenum Press, 1981.
- [2] HATHAWAY, R. J. – DAVENPORT, J. W. – BEZDEK, J. C.: Relational duals of the C-means clustering algorithms, *Pattern Recognition*, Vol. 22, No. 2, pp. 205–212, 1989.
- [3] CORSINI, P. – LAZZERINI, B. – MARCELLONI, F.: A new fuzzy relational clustering algorithm based on the fuzzy C-means algorithm, *Soft Computing*, Vol. 9, No. 6, pp. 439–447, 2005.
- [4] XIE, X. L. – BENI, G. A.: A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machines Intelligence*, Vol. PAMI-13, No. 8, pp. 841–847, 1991.
- [5] KRISHNAPURAM, R. – KELLER, J. M.: A possibilistic approach to clustering, *IEEE Transactions on Fuzzy Systems*, Vol. FS-1, No. 2, pp. 98–110, 1993.
- [6] VIATTCHENIN, D. A.: A new heuristic algorithm of fuzzy clustering, *Control and Cybernetics*, Vol. 33, No. 2, pp. 323–340, 2004.
- [7] VIATTCHENIN, D. A.: A direct algorithm of possibilistic clustering with partial supervision, *Journal of Automation, Mobile Robotics and Intelligent Systems*, Vol. 1, No. 3, pp. 29–38, 2007.
- [8] VIATTCHENIN, D. A.: Direct algorithms of fuzzy clustering based on the transitive closure operation and their application to outliers detection, *Artificial Intelligence*, No. 3, pp. 205–216, 2007. (in Russian)
- [9] VIATTCHENIN, D. A.: An outline for a heuristic approach to possibilistic clustering of the three-way data, *Journal of Uncertain Systems*, Vol. 3, No. 1, pp. 64–80, 2009.
- [10] VIATTCHENIN, D. A.: Analysis of the cluster structure robustness in non-stationary clustering problems, *Doklady BGUIR*, No. 6, pp. 91–98, 2009. (in Russian)
- [11] ICHINO, M. – YAGUCHI, H.: Generalized Minkowski metrics for mixed feature-type data analysis, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-24, No. 4, pp. 698–708.
- [12] KAUFMANN, A.: *Introduction to the Theory of Fuzzy Subsets*. New York: Academic Press, 1975.
- [13] VIATTCHENIN, D. A.: Validity measures for heuristic possibilistic clustering, *Information Technology and Control*, Vol. 39, No. 4, pp. 321–332, 2010.
- [14] JU, H. – YUAN, X.: Similarity measures on interval-valued fuzzy sets and application to pattern recognition, *Fuzzy Information and Engineering, D.Y. Cao (Ed.)*. Berlin: Springer-Verlag, 2007, pp. 875–883.
- [15] GRZEGORZEWSKI, P.: Distances between intuitionistic fuzzy sets and/or interval-valued fuzzy sets based on Hausdorff metric, *Fuzzy Sets and Systems*, Vol. 148, No. 2, pp. 319–328, 2004.
- [16] DUBOIS, D. – PRADE, H.: *Possibility Theory: An Approach to Computerized Processing of Uncertainty*. New York: Plenum Press, 1988.
- [17] VIATTCHENIN, D. A.: Derivation of fuzzy rules from interval-valued data, *International Journal of Computer Applications*, Vol. 7, No. 3, pp. 13–20, 2010.

Received February 21, 2011, accepted July 4, 2011

## BIOGRAPHY

**Dmitri A. Viattchenin** was born on 13.12.1972. In 1994 he graduated (MSc) at the Department of Mathematical Modelling and Data Analysis of the Faculty of Applied Mathematics and Informatics at Belarusian State University. He defended his PhD in the field of philosophy of sciences and technology in 1998. He is currently a senior researcher in the Laboratory of Information Protection. His research interests include hybrid techniques of possibilistic clustering and multivariate analysis, fuzzy control, simulation of complex systems, and philosophical aspects of artificial intelligence. He is the author of one book and over 115 papers in these areas.