Control and Cybernetics

vol. **47** (2018) No. 2

New developments in fuzzy clustering with emphasis on special types of tasks*

by

Dmitri A. Viattchenin

in collaboration with

Jan W. Owsiński and Janusz Kacprzyk

Systems Research Institute, Polish Academy of Sciences
Newelska 6, 01-447 Warszawa, Poland
{owsinski, kacprzyk}@ibspan.waw.pl

Abstract: The paper is devoted to a survey of work done in fuzzy clustering, mainly during the first decade of the 21^{st} century, and that with emphasis on various approaches to the problem, as well as various formulations of the very problem. That is why not only the classical formulations are treated, but several other problems, related to (the use of) clustering, like feature selection, inference systems, three-way clustering, and, on the other hand, such formulations of clustering as the possibilistic one or the one involving intuitionistic fuzzy sets. These are treated as the background for presentation of some specific ideas of the main author, concerning definite heuristic algorithms for effective solving of some of these problems.

Keywords: clustering, fuzzy sets, possibilistic clustering, inference, three-way clustering, feature selection

1. Introduction

Very soon after the inception of the fuzzy set theory at the end of the 1960s, due to Lotfi Zadeh (Zadeh, 1965), it turned out that the notions involved might be of practical, as well as methodological use in the area of clustering and associated tasks of data analysis. Thus, starting with Ruspini, approaches to clustering, based on application of the precepts of fuzzy set theory, have been developing very quickly (see, e.g., Ruspini, 1973). This was primarily due to James Bezdek, who worked on the development of this domain for several decades, and elaborated a complete set of approaches, which are still the basis of fuzzy-set-theory-founded clustering algorithms. This set of approaches referred explicitly to the k-means paradigm, introduced not so much earlier in its crisp form by MacQueen (even though prefigured much earlier, in the middle of the 1950s, by the Polish mathematician Hugo Steinhaus).

*Submitted by co-authors in September 2018; Accepted: November 2018.

The k-means type of fuzzy clustering proved to be so effective and efficient that it led to a host of more elaborate methodologies, among which, for instance, fuzzy-cluster-wise-modelling developed as an important branch in the domain.

In this short report we deal, however, primarily with some specialized fuzzy-set-based clustering methods, with specialization being understood here in two meanings: (1) the purpose of analysis is not just to produce partitions, composed of clusters (subsets), but, say, to identify classification rules or other structures, which may be demonstrated to have an affinity with appropriately defined clusters, or to group / analyse variables from the point of view of representation of the sample or population at hand; and (2) the methods applied do not belong to the current mainstream of fuzzy-set-based clustering methods, still essentially originating from the k-means motivated work of James Bezdek.

We shall be dealing in this short report, meant as an introduction to a much wider domain, separately with different kinds of data analysis related tasks and the methods of clustering that may be of use for these particular tasks.

2. Analysing the feature space with fuzzy clustering approaches

2.1. Introductory remarks

The reduction of dimensionality of the feature space analyzed is a very important problem in data analysis. Feature selection is meant here as the dimensionality reduction of the feature space of data that has initially contained a high number of features (high dimensionality of the initial data space). The purpose of the feature selection process is to choose a minimal number (subset) of the original set of features which still contain information that is essential for the discovering of relevant substructure(s) in the data, while reducing also the computational complexity, implied by using a high number of features in the source problem formulation. Feature selection has been a fertile field of research, and has been under intensive development since the 1970s, the resulting methods and algorithms mostly proving to be effective and efficient in removing irrelevant and/or redundant features, increasing the efficiency of learning (in the case of, say, machine learning-related problems), improving the learning performance, characterized by, for instance, predictive accuracy, and enhancing the comprehensibility of results obtained. (It is important to note that such methods may simplify and, indeed, make feasible, the applications of fuzzy sets in, for instance, control, see, e.g., Kacprzyk, 1997.) Thus, many different feature selection methods have been proposed, see, for example, Blum and Langley (1997), Kohavi and John (1997), Ghazavi and Liao (2008), or Dramiński et al. (2011).

Fuzzy clustering methods can well be applied to solve the problem of feature selection, and indeed, they are. So, in particular, a combination of feature selection with feature weights and semi-supervised fuzzy clustering in machine learning was proposed by Kong and Wang (2009). On the other hand, a fuzzy feature selection method based on clustering was proposed by Chitsaz, Taheri, and Katebi (2008). In the corresponding FACA-algorithm, each feature is assigned to different fuzzy clusters with different grades of membership. This comes from the basic underlying idea, standard for fuzzy clustering approaches, that each feature may belong not only to just one cluster, and it is much better to consider an association of each feature with other features in every cluster accounted for. Thus, precise relations between features are available during the selection of the most relevant features, as based on the co-membership in the obtained feature clusters.

An extension of the FACA-algorithm was then proposed in Chitsaz, Taheri, Katebi and Jahromi (2009), these authors having introduced four different techniques for implementing the stage of feature selection. In particular, for instance, by applying the chi-square test, their approach considers the dependence of each feature on class labels in the process of feature selection.

2.2. A heuristic approach to possibilistic clustering

The objective function based fuzzy clustering algorithms (predominantly, naturally, k-means-like) are the most widely employed methods in fuzzy clustering (see, for instance, Bezdek, Keller, Krishnapuram and Pal, 2005). There are also some heuristic clustering algorithms, which are based on the definition of the very concept of a cluster, and the purpose of these algorithms is to find clusters according to how they have been defined. Such algorithms are called direct classification (or clustering) algorithms (consult, e.g., Mandel, 1988).

An outline for a new heuristic method of fuzzy clustering was presented by Viattchenin (2004), who has considered a basic version of a direct clustering algorithm, while a version of such an algorithm, called the D-AFC(c)-algorithm, was then presented in Viattchenin (2007). The D-AFC(c)-algorithm can be considered as a direct possibilistic clustering algorithm, as this was demonstrated in Viattchenin (2007a). The D-AFC(c)-algorithm has been shown there to be a basis for the family of other heuristic possibilistic clustering algorithms.

The direct heuristic possibilistic clustering algorithms, originating from the basic idea, forwarded in Viattchenin (2004), can be divided into two types: relational and prototype-based. In particular, the family of direct relational heuristic possibilistic clustering algorithms includes:

- The D-AFC(c)-algorithm, which is based on the construction of an allotment (i.e. assignment of objects) among an a priori given number c of partially separate fuzzy clusters, see Viattchenin (2004);
- The D-AFC-PS(c)-algorithm, which is based on the construction of an allotment among an a priori given number c of partially separate fuzzy clusters in the presence of labeled objects, see Viattchenin (2007a);
- The D-PAFC-algorithm, which is based on the construction of an allotment among an *unknown* number of at least c fully separate fuzzy clusters, see Viattchenin (2009) for this version.

It should be noted that the D-PAFC-algorithm can be applied to solve the problem of informative feature selection. The corresponding method was also proposed in Viattchenin (2009).

On the other hand, the family of direct prototype-based heuristic possibilistic clustering algorithms, proposed by Viattchenin (2007b), includes:

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

The unique allotment among an unknown number ? of fuzzy clusters can be selected from the set of obtained allotments depending on the adopted tolerance threshold.

3. Fuzzy inference systems

3.1. An introduction

Fuzzy inference systems are presumably the best known and most popular applications of fuzzy logic and fuzzy sets theory. They can be employed to perform classification tasks, process simulation and diagnosis, online decision support and process control, just to name a few areas. So, the problem of generation of fuzzy classification rules (to be called fuzzy rules here, for brevity) is one of the most relevant problems in the development of fuzzy inference systems.

There are a number of approaches to learning fuzzy rules from data, they are, for instance, based on various techniques of evolutionary or neural computing, mostly aiming at the optimization of parameters of fuzzy rules. On the other hand, fuzzy clustering seems to be a very appealing and useful method for learning fuzzy rules since there is a close and canonical connection between fuzzy clusters and fuzzy rules. The idea of deriving fuzzy classification rules from data can be formulated as follows: the training data set is divided into homogeneous groups and a fuzzy rule is associated with (characterizes) each group.

Fuzzy clustering procedures are exactly pursuing this strategy: a fuzzy cluster is represented by a cluster center and the membership degree of a datum with respect to the cluster is decreasing with an increasing distance to the cluster center. So, each fuzzy rule of a fuzzy inference system can be characterized by a typical point and a membership function that is decreasing with an increasing distance to the typical point.

Let us consider some methods for extracting fuzzy rules from the data using fuzzy clustering algorithms. Some basic definitions should first be given.

Suppose that the training set contains n data pairs. Each pair is made up of an m_1-dimensional input vector and a c-dimensional output vector. We assume that the number of rules in the rule base of the fuzzy inference system is c. A Mamdani type (Mamdani and Assilian, 1975) rule l within the fuzzy inference system is written as follows:

$$If\ \hat{x}^1\ is\ B_l^1\ and\ \ldots and\ \hat{x}^{m_1}\ is\ B_l^{m_1}\ then\ y_1\ is\ C_1^l\ and \ldots and\ y_c\ is\ C_c^l, \quad (1)$$

where $B_l^{t_1}$, $t_1 \in \{1, \ldots, m_1\}$ and C_l^l, $l \in \{1, \ldots, c\}$ are fuzzy sets that define, respectively, the input and output space partitioning.

A fuzzy inference system, which is described by a set of fuzzy rules of the form (1) is a multiple input, multiple output (MIMO) system. Note that any fuzzy rule of the form (1) can be represented by c rules of the multiple input single output (MISO) type:

$$\begin{array}{l} If\ \hat{x}^1\ is\ B_l^1\ and\ \ldots and\ \hat{x}^{m_1}\ is\ B_l^{m_1}\ then\ y_1\ is\ C_1^l \\ \ldots \\ If\ \hat{x}^1\ is\ B_l^1\ and\ \ldots and\ \hat{x}^{m_1}\ is\ B_l^{m_1}\ then\ y_c\ is\ C_c^l \end{array}. \quad (2)$$

Let $B_l^{t_1}$ be characterized by its membership function $\gamma_{B_l^{t_1}}(\hat{x}^{t_1})$. These membership function can be of triangular, Gaussian, trapezoidal, or of any other suitable shape. Here, for the purpose of this report, we consider the trapezoidal and triangular membership functions, which are of a particular relevance for the real-life applications.

3.2. Classification rules obtained from fuzzy clustering

Fuzzy classification rules can be obtained directly from fuzzy clustering results. In general, a fuzzy clustering algorithm aims at minimizing the objective function, following the pattern proposed by James Bezdek (Bezdek, 1981):

$$Q(P, \bar{\mathrm{T}}) = \sum_{l=1}^{c} \sum_{i=1}^{n} \upsilon_{li}^{\gamma} d(x_i, \bar{\tau}^l), \quad (3)$$

subject to the constraints

$$\sum_{i=1}^{n} \upsilon_{li} > 1, \quad \forall i \in \{1, \ldots, n\}, \quad (4)$$

and

$$\sum_{l=1}^{c} \upsilon_{li} = 1, \quad \forall l \in \{1, \ldots, c\}, \quad (5)$$

where $X = \{x_1, \ldots, x_n\} \subseteq \Re^{m_1}$ is the analyzed data set, c is the number of fuzzy clusters A^l, $l = 1, \ldots, c$, in the fuzzy c-partition P, $\upsilon_{li} \in [0, 1]$ is the membership degree of object x_i with respect to the fuzzy cluster A^l, $\bar{\tau}^l \subseteq \Re^{m_1}$ is a prototype for a fuzzy cluster A^l, $d(x_i, \bar{\tau}^l)$ is the distance between the prototype $\bar{\tau}^l$ and the object x_i, and, finally, the parameter $\gamma > 1$ is an index of fuzziness. It is the selection of the concrete value of γ that determines whether the clusters in the partition tend to be more crisp or fuzzy. (As the notion of distance is omnipresent in these considerations, although they do not in any way depend upon any particular distance measure definition, let us only indicate the source for a variety of such definitions, namely Walesiak, 2002.)

The membership degrees can be calculated as following

$$\upsilon_{li} = \frac{1}{\sum\limits_{a=1}^{c} \left(\frac{d(x_i, \bar{\tau}^l)}{d(x_i, \bar{\tau}^a)} \right)^{1/(\gamma-1)}}, \tag{6}$$

and the prototypes can be obtained from the formula

$$\bar{\tau}^l = \frac{\sum\limits_{i=1}^{n} \upsilon_{li}^{\gamma} \cdot x_i}{\sum\limits_{i=1}^{n} \upsilon_{li}^{\gamma}}. \tag{7}$$

3.3. The possibilistic approach

The expressions (6) and (7) constitute, clearly, the necessary conditions for (3) to have a local minimum. However, condition (5) is, in general terms, hard to satisfy, mainly for reasons, related to numerical practicability, but also presents a constraint that is not always straightforwardly interpretable (is it certain that an object, or an observation, has to be precisely "distributed" among clusters, so that the sum of assignments is equal 1?). So, a possibilistic approach to clustering was proposed by Krishnapuram and Keller (1993). In particular, according to this possibilistic clustering approach, the objective function (3) is replaced by

$$Q(\Upsilon, \bar{\mathrm{T}}) = \sum_{l=1}^{c} \sum_{i=1}^{n} \left(\mu_{li}^{\psi} d(x_i, \bar{\tau}^l) + \eta_l (1 - \mu_{li})^{\psi} \right), \tag{8}$$

subject to a much more relaxed constraint, which, in fact corresponds to a possibilistic partition

$$\sum_{l=1}^{c} \mu_{li} > 1, \quad \forall l \in \{1, \ldots, c\}, \tag{9}$$

where c is the number of fuzzy clusters A^l, $l = 1, \ldots, c$, in the possibilistic partition Υ, $\mu_{li} \in [0, 1]$ is the set of values of possibilistic memberships of an i-th object, which are referred to here as typicality degrees, $\bar{\tau}^l \subseteq \Re^{m_1}$ is a prototype for the fuzzy cluster A^l, $d(x_i, \bar{\tau}^l)$ is the distance between the prototype $\bar{\tau}^l$ and the object x_i, and the parameter $\psi > 1$ has the interpretation analogous to that of the index of fuzziness.

The degrees of typicality can be calculated as follows

$$\mu_{li} = \frac{1}{1 + \left(d(x_i, \bar{\tau}^l)/\eta_l\right)^{1/(\psi-1)}}, \tag{10}$$

and the parameters η_l, $l = 1, \ldots, c$, are derived by

$$\eta_l = \frac{K}{\sum\limits_{i=1}^{n} \upsilon_{li}^{\psi}} \sum_{i=1}^{n} \upsilon_{li}^{\psi} d(x_i, \bar{\tau}^l), \tag{11}$$

where $K = 1$.

The principal idea of extracting fuzzy classification rules based on fuzzy clustering is as follows (see Höppner, Klawonn, Kruse and Runkler, 1999): each fuzzy cluster is assumed to be assigned to one class for classification and the membership degrees of the data to the clusters determine the degrees to which they can be classified as members of the corresponding classes. Thereby, with a fuzzy cluster that is assigned to a certain class, we can associate a linguistic rule, as this is often done with expressions, in which fuzzy components are used. The fuzzy cluster is projected into each single dimension leading to a fuzzy set being defined on the real line. From the mathematical point of view, the membership degree of the value $\hat{x}^{t_1}$ with respect to the t_1-th projection $\gamma_{B_l^{t_1}}(\hat{x}^{t_1})$ of the fuzzy cluster A^l, $l \in \{1, \ldots, c\}$ is the supremum over the membership degrees of all vectors with $\hat{x}^{t_1}$ as t_1-th component to the fuzzy cluster, i.e.

$$\gamma_{B_l^{t_1}}(\hat{x}^{t_1}) = \sup \left\{ 1 \Bigg/ \sum_{a=1}^{c} \left(d(x_i, \bar{\tau}^l)/d(x_i, \bar{\tau}^a)\right)^{1/(\gamma-1)} |x_i \right\} \tag{12}$$

where

$$x_i = (\hat{x}_i^1, \ldots, \hat{x}_i^{t_1-1}, \hat{x}_i^{t_1}, \hat{x}_i^{t_1+1}, \ldots, \hat{x}_i^{m_1}) \in \Re^{m_1},$$

or

$$\gamma_{B_l^{t_1}}(\hat{x}^{t_1}) = \sup \left\{ 1 \,/\, 1 + \left(d(x_i, \bar{\tau}^l)/\eta_l\right)^{1/(\gamma-1)} |x_i \right\} \tag{13}$$

where

$$x_i = (\hat{x}_i^1, \ldots, \hat{x}_i^{t_1-1}, \hat{x}_i^{t_1}, \hat{x}_i^{t_1+1}, \ldots, \hat{x}_i^{m_1}) \in \Re^{m_1},$$

with the latter expression being valid for the possibilistic case. An approximation of the fuzzy set by projecting only the data set and computing the convex hull of this projected fuzzy set, or approximating it by a trapezoidal or triangular membership function, is used for the rules obtained, conform to the reported proposal from Höppner, Klawonn, Kruse and Runkler (1999).

As mentioned already before, the objective function based fuzzy clustering algorithms are the most widespread methods in fuzzy clustering. Regarding those – the vast majority – based on the k-means principle, it must be admitted that they are highly effective and flexible, as well. However, they may be sensitive to the selection of an initial partition, and the fuzzy rules sought may depend on the selection of the concrete fuzzy clustering method employed. In particular, the well-known and effective algorithms, the GG (Gath-Geva) algorithm and the GK (Gustafsson-Kessel) algorithm of fuzzy clustering are recommended in Höppner, Klawonn, Kruse and Runkler (1999) for generation of fuzzy rules. This is quite natural, since the algorithms mentioned are oriented at identification of (fuzzy) cluster-wise models, and, in a general perspective, rules can be certainly interpreted as models. All the algorithms of possibilistic clustering are also the objective functions based algorithms.

On the other hand, there is the already mentioned heuristic approach to possibilistic clustering, which does not refer to an explicit objective function, outlined by Viattchenin (2004), and then further developed in following publications. Moreover, a method for an automatic generation of fuzzy inference systems using heuristic possibilistic clustering was outlined in Viattchenin (2010a). This method was thereafter extended for the case of the interval-valued data in Viattchenin (2010b).

4. Types of clustering structures and three-way data

4.1. Some introductory remarks

Most of the fuzzy clustering techniques are actually designed for handling crisp data, in a way augmented with their class membership degrees. However, the data can be (or, in fact, very often simply are) uncertain themselves. The initial data to be processed by clustering algorithms may be characterized by different types of uncertainty. For example, a brief review of uncertain data clustering methods is given in Viattchenin (2009). An interval uncertainty of the initial data can be considered to be the basic type of uncertainty in clustering.

The interval valued data can be considered as a particular case of the three-way data, in the sense of Sato and Sato (1994). The clustering problem for the case of the three-way data can be formulated as follows (see Sato and Sato, 1994, and Viattchenin, 2009). Let $X = \{x_1, ..., x_n\}$ be a set of objects, where objects are indexed by i, $i = 1, \ldots, n$; each object x_i being described by m_1 attributes, indexed by t_1, $t_1 = 1, \ldots, m_1$, so that an object x_i can be represented by a vector $x_i = (x_i^1, \ldots, x_i^{t_1}, \ldots, x_i^{m_1})$; each attribute $\hat{x}^{t_1}$, $t_1 = 1, \ldots, m_1$, can be characterized by m_2 values of binary attributes, so that $\hat{x}_i^{t_1} = (\hat{x}_i^{t_1(1)}, \ldots, \hat{x}_i^{t_1(t_2)}, \ldots, \hat{x}_i^{t_1(m_2)})$. For these notations, the three-way data can be represented as follows:

$$\hat{X}_{n \times m_1 \times m_2} = [\hat{x}_i^{t_1(t_2)}], \quad i = 1, \ldots, n, \quad t_1 = 1, \ldots, m_1, \quad t_2 = 1, \ldots, m_2. \quad (14)$$

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 $X = \{x_1, ..., x_n\}$ into c fuzzy clusters and the number of clusters c can be unknown, because it can depend on the concrete analyzed situation.

The initial data matrix (14) can be represented as a set of m_2 matrices $\hat{X}^{t_2}_{n\times m_1} = [\hat{x}^{t_1}_i]$, $i = 1,\ldots,n$, $t_1 = 1,\ldots,m_1$, corresponding to the "situations", and a "plausible" number c of fuzzy clusters can be different for each matrix $\hat{X}^{t_2}_{n\times m_1} = [\hat{x}^{t_1}_i]$, $t_2 \in \{1,\ldots,m_2\}$. The structure of clustering of the data set depends clearly on the type of the initial data.

Three types of the here pertinent clustering structures were distinguished in the paper of Viattchenin (2011). First, if the number of clusters c is constant for each matrix $\hat{X}^{t_2}_{n\times m_1} = [\hat{x}^{t_1}_i]$, $t_2 \in \{1,\ldots,m_2\}$, and the coordinates of prototypes $\{\bar{\tau}^1,\ldots,\bar{\tau}^?\}$ of the clusters $\{A^1,\ldots,A^?\}$ are constant, then the clustering structure is called *stable*. Second, if the current number of clusters c is constant for each matrix $\hat{X}^{t_2}_{n\times m_1} = [\hat{x}^{t_1}_i]$, $t_2 \in \{1,\ldots,m_2\}$, and the coordinates of prototypes of the clusters are not constant, then the clustering structure is called *quasi-stable*. Third, if the number of clusters c is different for the matrices $\hat{X}^{t_2}_{n\times m_1} = [\hat{x}^{t_1}_i]$, $t_2 = 1,\ldots,m_2$, then the clustering structure is called *unstable*.

Identification of the most plausible ("optimal") fuzzy clusters in the clustering structure sought for the uncertain data set X can be considered as a final goal of classification and the construction of the set of values of the most possible number of fuzzy clusters with their corresponding possibility degrees is an important step in this direction. The method of discovering a unique clustering structure, which corresponds to the most natural allocation of objects among fuzzy clusters for the uncertain data set was proposed by Viattchenin (2011). Following this, the idea of a novel approach to extracting fuzzy rules from the three-way data was presented by the same author. In this short introductory report we only concentrate on the basic essentials of the approach and the background methodological prerequisites. This outline is therefore now presented.

4.2. A novel approach to extracting fuzzy rules from the three-way data

Here we first discuss some basic concepts of the heuristic approach to possibilistic clustering. Then, we shall go over to the remarks on the preprocessing of the three-way data. Following this, we shall present a technique of extracting fuzzy rules from the three-way data.

4.2.1. Basic concepts of the heuristic method of possibilistic clustering

Heuristic algorithms of fuzzy clustering, such as considered, in particular, in Viattchenin (2004), are characterized by a low level of complexity and a high level of essential clarity. Some heuristic clustering algorithms are based on the definition of the concept of a cluster and the aim of these algorithms is to detect cluster that conform to a given definition. According to Mandel (1988), such algorithms can be called algorithms of direct classification or direct clustering algorithms.

As mentioned, an outline for a new heuristic method of fuzzy clustering was presented by Viattchenin (2004), where a basic version of a direct clustering algorithm was described A version of the algorithm that is called the D-AFC(c)-algorithm was given in Viattchenin (2007a). The D-AFC(c)-algorithm can be considered to constitute the direct algorithm of possibilistic clustering. This fact was demonstrated in Viattchenin (2007b). The D-AFC(c)-algorithm is the basis of an entire family of heuristic algorithms of possibilistic clustering. The heuristic approach to possibilistic clustering was further developed in other publications.

The direct heuristic algorithms of possibilistic clustering can be divided into two types: *relational* versus *prototype-based*. In particular, the family of direct relational heuristic algorithms of possibilistic clustering includes the variants that have been characterized in the present paper in Section 2.2., so that we shall not repeat this here.

On the other hand, the family of direct prototype-based clustering procedures, proposed in Viattchenin (2007b) includes:

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

Let us remind now some basic concepts of the heuristic method of possibilistic clustering in question. Thus, it is the concept of a fuzzy tolerance that is the basis for the concept of a fuzzy α-cluster. That is why the definition of a fuzzy tolerance must be considered in the first place.

Let $X = \{x_1, ..., x_n\}$ be an initial set of elements and $T : X \times X \to [0, 1]$ be 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 fuzzy tolerance is a fuzzy binary intransitive relation that is symmetric

$$\mu_T(x_i, x_j) = \mu_T(x_j, x_i), \quad \forall x_i, x_j \in X, \tag{15}$$

and reflexive

$$\mu_T(x_i, x_i) = 1, \quad \forall x_i \in X. \tag{16}$$

The notions of *powerful fuzzy tolerance*, *feeble fuzzy tolerance* and *strict feeble fuzzy tolerance* were considered in Viattchenin (2004) as well. In this context, the classical fuzzy tolerance in the sense of (15)–(16) has been referred to as the usual fuzzy tolerance in Viattchenin (2004). However, the essence of the method considered here does not depend on any particular kind of fuzzy tolerance, and is described for any fuzzy tolerance T.

Let α be an α-level value of the fuzzy tolerance T, $\alpha \in (0,1]$. Columns or rows of the fuzzy tolerance matrix are fuzzy sets $\{A^1, ..., A^n\}$ on X. Let A^l, $l \in \{1, \dots, n\}$, be a fuzzy set on X with $\mu_{A^l}(x_i) \in [0,1]$, $\forall x_i \in X$, being its membership function.

The α-level fuzzy set $A^l_{(\alpha)} = \{ (x_i, \mu_{A^l}(x_i)) | \mu_{A^l}(x_i) \geq \alpha, \ x_i \in X\}$ is a fuzzy α-cluster. So, $A^l_{(\alpha)} \subseteq A^l$, $\alpha \in (0,1]$, $A^l \in \{A^1, \dots, A^n\}$, and $\mu_{A^l}(x_i)$ is the membership degree of the element $x_i \in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1, \dots, n\}$. This membership degree will be denoted μ_{li} for brevity in further considerations. A value of α is the tolerance threshold of fuzzy α-cluster elements. The membership degree of an element $x_i \in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1, \dots, n\}$, can be defined as

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

where the α-level of a fuzzy set A^l ,$A^l_\alpha = \{x_i \in X | \mu_{A^l}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$, is the support of the fuzzy α-cluster $A^l_{(\alpha)}$.

The value of the membership function of each element of the fuzzy α-cluster is the degree of similarity of the object to some typical object of the fuzzy α-cluster. Moreover, the membership degree defines a possibility distribution function for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, and this possibility distribution function is here denoted $\pi_l(x_i)$.

Let $\{A^1_{(\alpha)}, ..., A^n_{(\alpha)}\}$ be the set of fuzzy α-clusters for some α. The point $\tau^l_e \in A^l_\alpha$, for which

$$\tau^l_e = \arg\max_{x_i} \mu_{li}, \quad \forall x_i \in A^l_\alpha \tag{18}$$

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

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

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

is met for all $A^l_{(\alpha)}$, $l = \overline{1,c}$, $c \leq n$, then this set is an allotment of elements of the set $X = \{x_1, ..., x_n\}$ among fuzzy α-clusters $\{A^l_{(\alpha)}, \ l = \overline{1,c}, 2 \leq c \leq n\}$ for some value of the tolerance threshold α. It should be noted that several allotments $R^\alpha_z(X)$ can exist for some tolerance threshold α. The number of allotments $R^\alpha_z(X)$ depend on the initial data structure. That is why the symbol z is introduced to denote the index of an allotment.

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

If some allotment $R^{\alpha}_z(X) = \{A^l_{(\alpha)} | l = 1, ..., n,\ c \leq n\}$ is considered appropriate for the problem considered, then this allotment is called an adequate allotment. In particular, if the conditions

$$\sum_{l=1}^{c} card(A^l_{\alpha}) \geq card(X),$$
$$\forall A^l_{(\alpha)} \in R^{\alpha}_z(X), \quad \alpha \in (0, 1], \quad card(R^{\alpha}_z(X)) = c, \tag{20}$$

and

$$card(A^l_{\alpha} \cap A^m_{\alpha}) \leq w, \quad \forall A^l_{(\alpha)}, A^m_{(\alpha)}, \quad l \neq m, \quad \alpha \in (0, 1], \tag{21}$$

are satisfied for all the fuzzy α-clusters $A^l_{(\alpha)}$, $l = 1, ..., n$, of some allotment $R^{\alpha}_z(X) = \{A^l_{(\alpha)} | l = 1, ..., n,\ c \leq n\}$, then this allotment is the allotment among particular separate fuzzy α-clusters and $w \in \{0, \ldots, n\}$ is the maximum number of elements in the intersection area of different fuzzy α-clusters. If $w = 0$ in the conditions (4.2) and (21), then this allotment is the allotment among fully separate fuzzy α-clusters.

An adequate allotment $R^{\alpha}_z(X)$ for some value of the tolerance threshold $\alpha \in (0, 1]$ is a family of fuzzy α-clusters, which are elements of the initial allotment $R^{\alpha}_I(X)$ for the value of α, and the family of fuzzy α-clusters satisfies the conditions (4.2) and (21). The problem consists in the selection of a unique adequate allotment $R^*(X)$ from the set B of adequate allotments, $B = \{R^{\alpha}_z(X)\}$, which is the class of possible solutions of the specific classification problem and $B = \{R^{\alpha}_z(X)\}$ depends on the parameters of the classification problem. In particular, the number c of fuzzy α-clusters is a parameter of the D-AFC(c)-algorithm.

The selection of the unique adequate allotment among a fixed number c of fuzzy α-clusters from the set $B = \{R^{\alpha}_z(X)\}$ of adequate allotments c is to be made on the basis of an evaluation of allotments. The criterion

$$F(R^{\alpha}_z(X), \alpha) = \sum_{l=1}^{c} \frac{1}{n_l} \sum_{i=1}^{n_l} \mu_{li} - \alpha \cdot c, \tag{22}$$

where c is the number of fuzzy α-clusters in the allotment $R^{\alpha}_z(X)$ and $n_l = card(A^l_{\alpha})$, $A^l_{(\alpha)} \in R^{\alpha}_z(X)$, is the number of elements in the support of the fuzzy α-cluster $A^l_{(\alpha)}$, can be used for evaluation of allotments.

The maximum value of the criterion (22) corresponds to the best allotment of objects among c fuzzy α-clusters. So, the classification problem can be formally characterized as the determination of a solution $R^*(X)$ satisfying

$$R^*(X) = \arg\max_{R_z^\alpha} (X) \in BF(R_z^\alpha(X), \alpha), \tag{23}$$

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

Thus, the problem of cluster analysis can be defined as the problem of discovering an allotment $R^*(X)$, resulting from the classification process, and the detection of a fixed number c of fuzzy α-clusters can be considered as the goal of classification. A description of the corresponding D-AFC(c)-algorithm is presented in the papers by Viattchenin (2004, 2007a, 2009, 2010b). These papers present also the examples of application, mainly on the well known data sets, such as the Iris data (see Anderson, 1935), or the data from Sato and Sato (1994).

The most "plausible" number c of fuzzy α-clusters in the allotment $R^*(X)$ sought can be considered as an index for the cluster validity problem for the D-AFC(c)-algorithm. Different validity measures for the D-AFC(c)-algorithm were proposed in Viattchenin (2010b). In particular, the measure of separation and compactness of the allotment can be defined in the following way:

$$V_{MSC}(R^*(X); c) = \frac{\sum\limits_{A^l_{(\alpha)} \in R^*(X)} D(A^l_{(\alpha)})}{c} + \frac{c}{n} \sum_{x_j \in \Theta} \mu_{lj} - \alpha, \tag{24}$$

where Θ is a set of elements x_j, $j \in \{1, \ldots, n\}$, in all of the intersection areas of different fuzzy α-clusters, and the density of a fuzzy α-cluster, $D(A^l_{(\alpha)})$, is defined in Viattchenin (2010b) as follows:

$$D(A^l_{(\alpha)}) = \frac{1}{n_l} \sum_{x_i \in A^l_\alpha} \mu_{li}, \tag{25}$$

where $n_l = card(A^l_\alpha)$, $A^l_{(\alpha)} \in R^*(X)$ and the membership degree μ_{li} is defined by the formula (17). The measure of separation and compactness of an allotment, $V_{MSC}(R^*(X); c)$, increases when c is closer to n. That is why the optimum value of c is obtained by minimizing $V_{MSC}(R^*(X); c)$ over $c = c_{\min}, \ldots, c_{\max}$, where $2 \leq c_{\min}$ and $c_{\max} < n$. So, the choice of the measure (24) can be interpreted as the search for an optimal number c of fuzzy α-clusters in the allotment $R^*(X)$ sought.

5. Concluding remarks

This short paper presented, against the background of the state of art in fuzzy clustering, with special emphasis on possibilistic version of this direction of research, and in the context of some specialized tasks (rule extraction, three-way data classification, etc.), the place and role of the work, primarily done by Dmitri

A. Viattchenin, on the heuristic methods of clustering. These methods, which start from very basic precepts, concerning what the fuzzy / possibilistic partition (allotment) is – and therefore what is the meaning of the clusters, forming this partition – can be effectively used for solving the tasks here mentioned. A separate issue is constituted by the numerical questions, accompanying the realization of the respective algorithms, and this ought to be considered in detail in future studies, conducted by the associates of Dmitri Viattchenin.

References

Anderson E. (1935) The irises of the Gaspe Peninsula. *Bulletin of the American Iris Society*, **59** (1): 2-5.

Bezdek J. C. (1981) *Pattern Recognition with Fuzzy Objective Function Algorithms.* Plenum Press, New York.

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

Blum A. and Langley P. (1997) Selection of relevant features and examples in machine learning. *Artificial Intelligence*, **97** (1-2), 245-271.

Chaudhuri B. B. and Bhattacharya A. (2001) On correlation between two fuzzy sets. *Fuzzy Sets and Systems* **118** (3), 447-456.

Chiang D.-A. and Lin N. P. (1999) Correlation of fuzzy sets. *Fuzzy Sets and Systems* **102** (2), 221-226.

Chitsaz E., Taheri M. and Katebi S. D. (2008) A fuzzy approach to clustering and selecting features for classification of gene expression data. *Proc. World Congress of Engineering (WCE'2008)*, International Asssociation of Engineers, 1650-1655
http://www.iaeng.org/publication/WCE2008/

Chitsaz E., Taheri M., Katebi S. D. and Jahromi M. Z. (2009) An improved fuzzy feature clustering and selection based on chi-squared-test. *Proc. Int. Multiconference of Engineers and Computer Scientists (IMECS'2009).* International Association of Engineers, 35-40
http://www. iaeng.org/publication/WCE2008/

Dramiński M., Kierczak M., Nowak-Brzezińska A., Koronacki J. and Komorowski J. (2011) The Monte Carlo feature selection and interdependency discovery is unbiased. *Control and Cybernetics*, **40** (2), 199-211.

Ghazavi S. N. and Liao T. W. (2008) Medical data mining by fuzzy modeling with selected features. *Artificial Intelligence in Medicine*, **43** (3), 195-206.

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

Kacprzyk J. (1997) *Multistage Fuzzy Control.* Wiley, Chichester.

Kohavi R. and John G. (1997) Wrappers for feature subset selection. *Artificial Intelligence*, **97** (1-2), 273-324.

Kong Y.-Q. and Wang S.-T. (2009) Feature selection and semi-supervised fuzzy clustering. *Fuzzy Information and Engineering* **1** (2), 179-190.

Krishnapuram R. and Keller J. M. (1993) A possibilistic approach to clustering. *IEEE Trans. on Fuzzy Systems*, **1** (2), 98-110.

Mamdani E. H. and Assilian S. (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies* 7, 1-13.

Mandel I. D. (1988) *Clustering Analysis.* Finansy i Statistika, Moscow (in Russian).

Murthy C. A., Pal S. K. and Dutta Majumder D. (1985) Correlation between two fuzzy membership functions. *Fuzzy Sets and Systems* **17** (1), 23-38.

Ruspini E. H. (1973) New experimental results in fuzzy clustering. *Information Sciences*, **6**, 273-284.

Sato M. and Sato Y. (1994) On a multicriteria fuzzy clustering method for 3-way data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 2, 127-142.

Sato-Ilic M. and L.C. Jain L. C. (2006) *Innovations in Fuzzy Clustering: Theory and Applications.* Springer-Verlag, Heidelberg.

Viattchenin D. A. (2004) A new heuristic algorithm of fuzzy clustering. *Control and Cybernetics* **33** (2), 323-340.

Viattchenin D. A. (2007a) A direct algorithm of possibilistic clustering with partial supervision. *Journal of Automation, Mobile Robotics and Intelligent Systems* **1** (3), 29-38.

Viattchenin D. A. (2007b) Direct algorithms of fuzzy clustering based on the transitive closure operation and their application to outliers detection. *Artificial Intelligence* 3, 205-216 (in Russian).

Viattchenin D. A. (2009a) An algorithm for detecting the principal allotment among fuzzy clusters and its application as a technique of reduction of analyzed features space dimensionality. *J. Information and Organizational Sciences* **33** (1), 205-217.

Viattchenin D. A. (2009b) An outline for a heuristic approach to possibilistic clustering of the three-way data. *Journal of Uncertain Systems* 3, 64-80.

Viattchenin D. A. (2010a) Automatic generation of fuzzy inference systems using heuristic possibilistic clustering. *Journal of Automation, Mobile Robotics and Intelligent Systems* **4**(3) 36-44.

Viattchenin D. A. (2010b) Derivation of fuzzy rules from interval-valued data. *International Journal of Computer Applications* **7**(3) 13-20.

Viattchenin D. A. (2010c) Validity measures for heuristic possibilistic clustering. *Information Technology and Control* 39, 321-332.

Viattchenin D. A. (2011) Constructing stable clustering structure for uncertain data set. *Acta Electrotechnica et Informatica* **11**(3) 42-50.

Walesiak M. (2002) *A generalized distance measure in statistical multivarate analysis.*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław (in Polish).

Zadeh L. (1965) Fuzzy sets. *Information and Control*; 8, 338–353.