Extracting Fuzzy Classifications Rules from Three-way Data

Submitted 28th October 2013; accepted 19th January 2013

Janusz Kacprzyk, Jan W. Owsinski, Dmitri A. Viattchenin

DOI: 10.14313/JAMRIS_2-2014/19

Abstract: *The paper deals in the conceptual way with the problem of extracting fuzzy classification rules from the three-way data meant in the sense of Sato and Sato [7]. A novel technique based on a heuristic method of possibilistic clustering is proposed. A description of basic concepts of a heuristic method of possibilistic clustering based on concept of an allotment is provided. A preprocessing technique for three-way data is shown. An extended method of constructing fuzzy classification rules based on clustering results is proposed. An illustrative example of the method's application to the Sato and Sato's data [7] is provided. An analysis of the experimental results obtained with some conclusions are given.*

Keywords: *three-way data, possibilistic clustering, fuzzy cluster, typical point, fuzzy rule*

1. Introduction

Some remarks on fuzzy inference systems and a brief review of methods of extracting fuzzy classification rules based on fuzzy clustering are considered in the first subsection. The second subsection includes a discussion of basic types of the uncertainty of the data and specifies the purpose of the paper.

1.1. Preliminary Remarks

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, 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, 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 (characterizing) 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 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 a m_1 – dimensional input vector and a $\tilde{n}$-dimensional output vector. We assume that the number of rules in the rule base of the fuzzy inference system is $\tilde{n}$. A Mamdani type [5] rule l within the fuzzy inference system is written as follows:

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

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 an 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 inputs, multiple outputs (MIMO) system. Note that any fuzzy rule of the form (1) can be presented by c rules of the multiple inputs single output (MISO) type:

$$\begin{aligned} &If\ \hat{x}^1\ is\ B_l^1\ and \ldots and\ \hat{x}^{m1}\ is\ B_l^{m1}\ then\ y_1\ is\ C_1^l \\ &\ldots \\ &If\ \hat{x}^1\ is\ B_l^1\ and \ldots and\ \hat{x}^{m1}\ is\ B_l^{m1}\ then\ y_c\ is\ C_c^l. \end{aligned} \qquad (2)$$

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

Fuzzy classification rules can be obtained directly from fuzzy clustering results. In general, a fuzzy clustering algorithm aims at minimizing the objective function [1]

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

subject to the constraints

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

and

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

where $X=\{x_1,\ldots,x_n\}\subseteq\Re^{m_1}$ is the 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 to fuzzy cluster A^l, $\overline{\tau}^l\subseteq\Re^{m_1}$ is a prototype for a fuzzy cluster A^l, $d(x_i,\overline{\tau}^l)$ is a distance between the prototype $\overline{\tau}^l$ and object x_i, and the parameter $\gamma>1$ is an index of fuzziness. The selection of the value of γ determines whether the cluster tend to be more crisp or fuzzy.

The membership degrees can be calculated as following

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

and the prototypes can be obtained from the formula

$$\overline{\tau}^l=\frac{\sum_{i=1}^{n}\upsilon_i^{\gamma}\cdot x_i}{\sum_{i=1}^{n}\upsilon_i^{\gamma}}. \tag{7}$$

The expressions (6) and (7) are clearly the necessary conditions for (3) to have a local minimum. However, the condition (5) is hard to satisfy for obvious reasons. So, a possibilistic approach to clustering was proposed by Krishnapuram and Keller [4]. In particular, the objective function (3) is replaced by

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

subject to the constraint of a possibilistic partition

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

where c is the number of fuzzy clusters A^l, $l=1,\ldots,c$, in the possibilistic partition Y, $\mu_{li}\in[0,1]$ is the possibilistic memberships which are typicality degrees, $\overline{\tau}^l\subseteq\Re^{m_1}$ is a prototype for the fuzzy cluster A^l, $d(x_i,\overline{\tau}^l)$ is a distance between the prototype $\overline{\tau}^l$ and object x_i, and the parameter $\psi>1$ is meant analogously as the index of fuzziness.

The degrees of typicality can be calculated as follows:

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

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

$$\eta_l=\frac{K}{\sum_{i=1}^{n}\upsilon_i^{\psi}}\sum_{i=1}^{n}\upsilon_i^{\psi}d(x_i,\tau^l), \tag{11}$$

where $K=1$.

The principal idea of extracting fuzzy classification rules based on fuzzy clustering is as follows (cf. Höppner, Klawonn, Kruse and Runkler [2]). 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. So, with a fuzzy cluster that is assigned to the some class we can associate a linguistic rule. The fuzzy cluster is projected into each single dimension leading to a fuzzy set defined in the real line. From a mathematical point of view, the membership degree of the value $\hat{x}^{t_1}$ 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.

$$\gamma_{B_l^{t_1}}(\hat{x}^{t_1})=\sup\left\{1\Big/\sum_{a=1}^{c}\left(d(x_i,\tau^l)/d(x_i,\tau^a)\right)^{1/(\gamma-1)}\,\middle|\,x_i=(\hat{x}_i^1,\ldots,\hat{x}_i^{t_1-1},\hat{x}_i^{t_1},\hat{x}_i^{t_1+1},\ldots,\hat{x}_i^{m_1})\in\Re^{m_1}\right\} \tag{12}$$

or

$$\gamma_{B_l^{t_1}}(\hat{x}^{t_1})=\sup\left\{1\Big/1+\left(d(x_i,\tau^l)/\eta_l\right)^{1/(\psi-1)}\,\middle|\,x_i=(\hat{x}_i^1,\ldots,\hat{x}_i^{t_1-1},\hat{x}_i^{t_1},\hat{x}_i^{t_1+1},\ldots,\hat{x}_i^{m_1})\in\Re^{m_1}\right\} \tag{13}$$

in 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 [2].

The objective function based fuzzy clustering algorithms are the most widespread methods in fuzzy clustering. However, they may be sensitive to the selection of an initial partition, and the fuzzy rules sought may depend on the selection of the fuzzy clustering method employed. In particular, the GG (Gath-Geva) algorithm and the GK (Gustafsson-Kessel) algorithms of fuzzy clustering are recommended in [2] for the generation of fuzzy rules. All algorithms of possibilistic clustering are also the objective functions based algorithms.

On the other hand, a heuristic approach to possibilistic clustering was outlined by Viattchenin [8] and then further developed in next publications. Moreover, a method for an automatic generation of fuzzy inference systems using heuristic possibilistic clustering was outlined by Viattchenin [13]. This method was extended for the case of the interval-valued data in [14].

1.2. Types of Clustering Structures

Most fuzzy clustering techniques are designed for handling crisp data augmented with their class membership degrees. However, the data can be uncertain. The initial data to be processed by clustering algorithms may be characterized y different types of uncertainty. For example, a brief review of uncertain data clustering methods is given in [11]. An interval uncertainty of the initial data is a basic type of uncertainty in clustering.

The interval valued data are a particular case of the three-way data as meant by Sato and Sato [7]. The clustering problem for the case of the three-way data can be formulated as follows [7, 11]. 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 is 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 $x_i=(x_i^1,\ldots,x_i^{t1},\ldots,x_i^{m1})$; each attribute $\hat{x}^{t1}$, $t_1=1,\ldots,m_1$, can be characterized by m_2 values of binary attributes, so that $\hat{x}_i^{t1}=(\hat{x}_i^{t1(1)},\ldots,\hat{x}_i^{t1(t2)},\ldots,\hat{x}_i^{t1(m2)})$. Therefore, the three-way data can be represented as follows:

$$\hat{X}_{n\times m1\times m2}=[\hat{x}_i^{t1(t2)}],\ i=1,\ldots,n,$$
$$t_1=1,\ldots,m_1,\ t_2=1,\ldots,m_2. \tag{14}$$

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 situation.

The initial data matrix (14) can be represented as a set of m_2 matrices $\hat{X}^{t2}_{n\times m1}=[\hat{x}_i^{t1}]$, $i=1,\ldots,n$, $t_1=1,\ldots,m_1$ and a "plausible" number c of fuzzy clusters can be different for each matrix $\hat{X}^{t2}_{n\times m1}=[\hat{x}_i^{t1}]$, $t_2\in\{1,\ldots,m_2\}$. The structure of clustering of the data set depends clearly on the type of the initial data.

Three types of the clustering structures were defined by Viattchenin [16]. First, if the number of clusters c is constant for each matrix $\hat{X}^{t2}_{n\times m1}=[\hat{x}_i^{t1}]$, $t_2\in\{1,\ldots,m_2\}$, and the coordinates of prototypes $\{\overline{\tau}^1,\ldots,\overline{\tau}^{\tilde{n}}\}$ of the clusters $\{A^1,\ldots,A^{\tilde{n}}\}$ are constant, then the clustering structure is called stable. Second, if the current number of clusters c is constant for each matrix $\hat{X}^{t2}_{n\times m1}=[\hat{x}_i^{t1}]$, $t_2\in\{1,\ldots,m_2\}$, 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 $\hat{X}^{t2}_{n\times m1}=[\hat{x}_i^{t1}]$, $t_2=1,\ldots,m_2$, then the clustering structure is called unstable.

The detection of the most plausible 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 way. The method of discovering a unique clustering structure which corresponds to most natural allocation of objects among fuzzy clusters for the uncertain data set was proposed by Viattchenin [16].

The main goal of this paper is to present the idea of a novel approach to extracting fuzzy rules from the three-way data. The contents of this paper is as follows: in the second section the basic concepts of the heuristic approach to possibilistic clustering are presented, a preprocessing technique for the three-way data is given and a technique of prototyping fuzzy classification rules from the three-way data based on the heuristic possibilistic clustering is proposed. In the third section an illustrative example of application of the proposed technique to Sato and Sato's [7] three-way data set is given, and in the fourth section some conclusions are formulated.

2. A Novel Approach to Extracting Fuzzy Rules from the Three-way Data

In the first subsection some basic concepts of the heuristic approach to possibilistic clustering are discussed. The second subsection includes some remarks on the preprocessing of the three-way data. A technique of extracting fuzzy rules from the three-way data is described in the third subsection.

2.1. Basic Concepts of the Heuristic Method of Possibilistic Clustering

Heuristic algorithms of fuzzy clustering 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. Due to Mandel [6] such algorithms can be called algorithms of direct classification or direct clustering algorithms.

An outline for a new heuristic method of fuzzy clustering was presented by Viattchenin [8] 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 [9]. The D-AFC(c)-algorithm can be considered as a direct algorithm of possibilistic clustering. This fact was demonstrated in [9]. 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 D-AFC(c)-algorithm via the construction of an allotment among an a priori given number *c* of partially separate fuzzy clusters [8];

The D-AFC-PS(c)-algorithm via the construction of an allotment among an a priori given number *c* of partially separate fuzzy clusters in the presence of labeled objects [9];

The D-PAFC-algorithm via the construction of an allotment among an unknown number of at least *c* fully separate fuzzy clusters [12].

On the other hand, the family of direct prototype-based clustering procedures, proposed in [10] includes:

The D-AFC-TC-algorithm via the construction of an allotment among an unknown number *c* of fully separate fuzzy clusters;

The D-PAFC-TC-algorithm 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 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 some basic concepts of the heuristic method of possibilistic clustering in question. The concept of a fuzzy tolerance 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

$$\mu_T(x_i,x_j)=\mu_T(x_j,x_i),\ \forall x_i,x_j\in X, \qquad (15)$$

and reflexive

$$\mu_T(x_i,x_i)=1,\ \forall x_i\in X. \qquad (16)$$

The notions of a powerful fuzzy tolerance, a feeble fuzzy tolerance and a strict feeble fuzzy tolerance were considered by Viattchenin [8] as well. In this context the classical fuzzy tolerance in the sense of (15)–(16) has been called an usual fuzzy tolerance in [8]. However, the essence of the method considered here does not depend on any particular kind of a 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, $l\in\{1,\ldots,n\}$, 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, $A^l_{(\alpha)}\subseteq A^l$, $\alpha\in[0,1]$, $A^l\in\{A^1,\ldots,A^n\}$, and $\mu_{A^l}(x_i)$ is the membership degree of the element $x_i\in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha\in[0,1]$, $l\in\{1,\ldots,n\}$. This membership degree will be denoted μ_{li} for brevity in further considerations. A value of α is a tolerance threshold of fuzzy α-cluster elements. The membership degree of an element $x_i\in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha\in[0,1]$, $l\in\{1,\ldots,n\}$, can be defined as

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

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 denoted $\pi_l(x_i)$.

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

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

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 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

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

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 the index of an allotment.

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 $R_z^\alpha(X)=\{A_{(\alpha)}^l \mid l=1,...,n, c\le n\}$ is considered appropriate for the problem considered, then this allotment is called an adequate allotment. In particular, if the conditions

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

$$\alpha\in(0,1],\ card(R_z^\alpha(X))=c, \quad (20)$$

and

$$card(A_\alpha^l \cap A_\alpha^m) \le w,\ \forall A_{(\alpha)}^l, A_{(\alpha)}^m,$$

$$l\ne m,\ \alpha\in(0,1], \quad (21)$$

are met for all fuzzy α-clusters $A_{(\alpha)}^l$, $l=1,...,n$, of some allotment $R_z^\alpha(X)=\{A_{(\alpha)}^l \mid l=1,...,n, c\le n\}$, 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 (20) and (21), then this allotment is the allotment among fully separate fuzzy α-clusters.

An adequate allotment $R_z^\alpha(X)$ 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 $R_l^\alpha(X)$ for the value of α, and the family of fuzzy α-clusters satisfies the conditions (20) and (21). The problem consists in the selection of an 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 specific classification problem and $B=\{R_z^\alpha(X)\}$ 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 $B=\{R_z^\alpha(X)\}$ of adequate allotments c is to be made on the basis of an evaluation of allotments. The criterion

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

where c is the number of fuzzy α-clusters in the allotment $R_z^\alpha(X)$ and $n_l=card(A_\alpha^l)$, $A_{(\alpha)}^l\in R_z^\alpha(X)$, is the number of elements in the support of the fuzzy α-cluster $A_{(\alpha)}^l$, 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

$$R^*(X)=\arg\max_{R_z^\alpha(X)\in B} F(R_z^\alpha(X),\alpha), \quad (23)$$

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 [8, 9, 11, 15].

The most "plausible" number $\tilde{n}$ 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 [15]. In particular, the measure of separation and compactness of the allotment can be defined in the following way:

$$V_{MSC}(R^*(X);c)=\frac{\sum_{A_{(\alpha)}^l\in R^*(X)} D(A_{(\alpha)}^l)}{c}+\frac{c}{n}\sum_{x_i\in\Theta}\mu_{lj}-\alpha, \quad (24)$$

where Θ is a set of elements x_j, $j\in\{1,\ldots,n\}$, in all intersection areas of different fuzzy α-clusters, and the density of fa uzzy α-cluster, $D(A_{(\alpha)}^l)$, is defined in [15] as follows:

$$D(A_{(\alpha)}^l)=\frac{1}{n_l}\sum_{x_i\in A_\alpha^l}\mu_{li}, \quad (25)$$

where $n_l=card(A_\alpha^l)$, $A_{(\alpha)}^l\in R^*(X)$ and 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 $2\le c_{\min}$ and $c_{\max}<n$. So, the choice of the measure (24) can be interpreted as the search for an optimal number $\tilde{n}$ of fuzzy α-clusters in the allotment $R^*(X)$ sought.

2.2. Remarks on the Preprocessing of Three-way Data

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

$$x_i^{t1(t2)}=\frac{\hat{x}_i^{t1(t2)}-\min_{i,t2}\hat{x}_i^{t1(t2)}}{\max_{i,t2}\hat{x}_i^{t1(t2)}-\min_{i,t2}\hat{x}_i^{t1(t2)}}. \quad (26)$$

So, each object x_i, $i=1,\ldots,n$, from the initial set $X=\{x_1,\ldots,x_n\}$ can be considered as a type-two fuzzy set and $x_i^{t1(t2)}=\mu_{x_i}(x^{t1(t2)})$, $i=1,\ldots,n$; $t_1=1,\ldots,m_1$, $t_2=1,\ldots,m_2$, $x^{t1(t2)}=\mu_{t_1}(x^{t2})\in[0,1]$, $t_1=1,\ldots,m_1$, $t_2=1,\ldots,m_2$,are its membership functions. In the case of three-way data each object x_i, $i=1,\ldots,n$ can be represented as a matrix $X_{(i)m_1\times m_2}=[x_i^{t1(t2)}]$, $t_1=1,\ldots,m_1$, $t_2=1,\ldots,m_2$.

Dissimilarity coefficients between the objects can be constructed on the basis of generalizations of distances between fuzzy sets [11] and these generalizations take into account dissimilarities between the attributes of objects and the situations. In particular, a generalization of the squared normalized Euclidean distance for type-two fuzzy sets can be described by

$$\varepsilon_G(x_i, x_j) = \frac{1}{m_1}\sum_{t1=1}^{m1}\left(\frac{1}{m_2^2}\sum_{u1,v1=1}^{m2}\left(\mu_{x_i}(x^{t1(u1)}) - \mu_{x_j}(x^{t1(v1)})\right)^2\right), \quad (27)$$

for all $i, j = 1,\ldots,n$. In the case of $m_2 = 1$, the formula (27) can be rewritten as the usual squared normalized Euclidean distance [3]:

$$\varepsilon(x_i, x_j) = \frac{1}{m_1}\sum_{t1=1}^{m1}\left(\mu_{x_i}(x^{t1}) - \mu_{x_j}(x^{t1})\right)^2, \quad i, j = 1,\ldots,n. \quad (28)$$

The fuzzy tolerance matrix $I = [\mu_I(x_i, x_j)]$, $i, j = 1,\ldots,n$, can be obtained by the application of the complement operation

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j), \ \forall i, j = 1,\ldots,n, \quad (29)$$

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

However, the value m_2 can be different for different attributes $\hat{x}^{t1}$, $t_1 \in \{1,\ldots,m_1\}$, or the value m_2 of grades for a fixed attribute $\hat{x}^{t1}$, $t_1 \in \{1,\ldots,m_1\}$ can be different for different objects x_i, $i \in \{1,\ldots,n\}$. So, each object x_i, $i = 1,\ldots,n$, cannot be presented as a matrix $X_{(i)m_1 \times m_2} = [x_i^{t1(t2)}]$, $t_1 = 1,\ldots,m_1$, $t_2 = 1,\ldots,m_2$, because the value m_2, which is general for all attributes $\hat{x}^{t1}$, $t_1 \in \{1,\ldots,m_1\}$, must be established. In these cases a general value m_2 can be defined as follows:

$$m_2 = \max_{t1} m_2^{(t1)}, \ t_1 = 1,\ldots,m_1, \quad (30)$$

where the number of grades of each attribute $\hat{x}^{t1}$, $t_1 \in \{1,\ldots,m_1\}$, is denoted by $m_2^{(t1)}$. However, values $x_i^{t1(t2)}$, $i \in \{1,\ldots,n\}$ may be unknown for some objects $x_i \in X$, $i \in \{1,\ldots,n\}$. In such a case, an unknown values $x_i^{t1(t2)}$, $i \in \{1,\ldots,n\}$, can be defined heuristically as follows:

$$x_i^{t1(t2)} = \max_{t1} t_2^{(t1)}, \ i \in \{1,\ldots,n\}, \ t_2 = 1,\ldots,m_2^{(t1)}. \quad (31)$$

Obviously, the preprocessing method for the three-way data can be very simply generalized for the case of p-way data, for p >3.

2.3. A Method of Fuzzy Rules Extraction from the Three-way Data

Let us consider a method of extracting fuzzy classification rules based on a heuristic method of possibilistic clustering [13]. In the following, we will consider that the Mamdani type fuzzy inference system is a multiple inputs, multiple outputs system (MIMO).

The antecedent of a fuzzy rule in the fuzzy inference system defines a decision region in the m_1-dimensional feature space. Let us consider a fuzzy rule (1) where $B_l^{t_1}$, $t_1 = 1,\ldots,m_1$, $l \in \{1,\ldots,c\}$, is a fuzzy

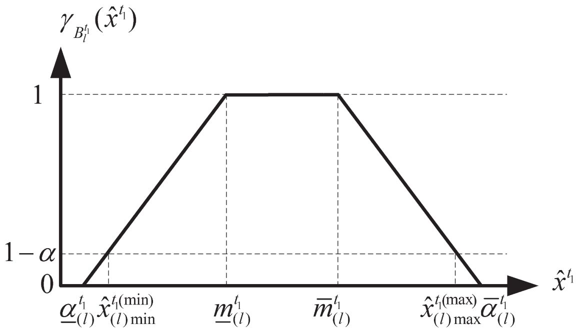

Fig. 1. A trapezoidal membership function for an antecedent fuzzy set

set associate with the attribute variable $\hat{x}^{t_1}$. Let $B_l^{t_1}$ be characterized by its trapezoidal membership function $\gamma_{B_l^{t1}}(\hat{x}^{t_1})$ which is presented in Fig. 1.

So, the fuzzy set B_l^{t1} can be defined by four parameters, $B_l^{t1} = (\underline{a}_{(l)}^{t1}, \underline{m}_{(l)}^{t1}, \overline{m}_{(l)}^{t1}, \overline{a}_{(l)}^{t1})$. A triangular fuzzy set $B_l^{t_1} = (\underline{a}_{(l)}^{t_1}, m_{(l)}^{t_1}, \overline{a}_{(l)}^{t_1})$ can be considered as a particular case of the trapezoidal fuzzy set where $\underline{m}_{(l)}^{t_1} = \overline{m}_{(l)}^{t_1}$. The idea of deriving fuzzy rules from fuzzy α-clusters was outlined by Viattchenin [13] and this method can be extended to the case of the three-way data as follows. We apply the D-AFC(c)-algorithm to the given three-way data and then obtain for each fuzzy α-cluster $A_{(\alpha)}^l$, $l \in \{1,\ldots,c\}$ a kernel $K(A_{(\alpha)}^l)$ and a support A_α^l. The value of the tolerance threshold $\alpha \in (0,1]$, which corresponds to an allotment $R^*(X) = \{A_{(\alpha)}^1,\ldots,A_{(\alpha)}^c\}$, is an additional result of classification. The situation of the three-way data can be described by the expression $\hat{x}_i^{t1} = (\hat{x}_i^{t1(\min)}, \hat{x}_i^{t1(\max)})$, $t_1 = 1,\ldots,m_1$, $i = 1,\ldots,n$. In particular, the interval $[\hat{x}_{(l)\min}^{t1(\min)}, \hat{x}_{(l)\max}^{t1(\min)}]$ of values of each attribute $\hat{x}^{t1} = (\hat{x}^{t1(\min)}, \hat{x}^{t1(\max)})$, $t_1 \in \{1,\ldots,m_1\}$ for the support A_α^l should be calculated. We calculate the interval $[\hat{x}_{(l)\min}^{t1(\min)}, \hat{x}_{(l)\max}^{t1(\min)}]$ of values of each attribute $\hat{x}^{t1}$, $t_1 \in \{1,\ldots,m_1\}$, for the support A_α^l. The value $\hat{x}_{(l)\min}^{t1(\min)}$ can be obtained ass

$$\hat{x}_{(l)\min}^{t1(\min)} = \min_{x_i \in A_\alpha^l} \hat{x}^{t1(\min)},$$

$$\forall t_1 \in \{1,\ldots,m_1\}, \ \forall l \in \{1,\ldots,c\}, \quad (32)$$

and the value $\hat{x}_{(l)\max}^{t1(\max)}$, $t_1 \in \{1,\ldots,m_1\}$, as

$$\hat{x}_{(l)\max}^{t1(\max)} = \max_{x_i \in A_\alpha^l} \hat{x}^{t1(\max)},$$

$$\forall t_1 \in \{1,\ldots,m\}, \ \forall l \in \{1,\ldots,c\}. \quad (33)$$

The parameter $\underline{a}_{(l)}^{t1}$ can be obtained from

$$\gamma_{B_l^{t1}}(\hat{x}_{(l)\min}^{t1(\min)}) = (1-\alpha), \ \gamma_{B_l^{t1}}(\underline{a}_{(l)}^{t1}) = 0, \quad (34)$$

and the parameter $\overline{a}^{t1}_{(l)}$ from

$$\gamma_{B_l^{t1}}(\hat{x}^{t1(\max)}_{(l)\max}) = (1-\alpha)\,,\ \gamma_{B_l^{t1}}(\overline{a}^{t1}_{(l)}) = 0\,. \quad (35)$$

We calculate the value $\underline{\hat{x}}^{t1(\min)}_{(l)}$ for all typical points $\tau^l_e \in K(A^l_{(\alpha)})$ of the fuzzy α-cluster $A^l_{(\alpha)}$, $l \in \{1,\ldots,c\}$, as follows:

$$\underline{\hat{x}}^{t1(\min)}_{(l)} = \min_{\tau^l_e \in K(A^l_{(\alpha)})} \hat{x}^{t1(\min)},\ \forall e \in \{1,\ldots,|l|\}\,, \quad (36)$$

and the value $\hat{\overline{x}}^{t1(\max)}_{(l)}$ can be obtained from

$$\hat{\overline{x}}^{t1(\max)}_{(l)} = \max_{\tau^l_e \in K(A^l_{(\alpha)})} \hat{x}^{t1(\max)},\ \forall e \in \{1,\ldots,|l|\}\,. \quad (37)$$

Thus, the parameter $\underline{m}^{t1}_{(l)}$ can be calculated from

$$\gamma_{B_l^{t1}}(\underline{\hat{x}}^{t1(\min)}_{(l)}) = \gamma_{B_l^{t1}}(\underline{m}^{t1}_{(l)}) = 1\,, \quad (38)$$

and the parameter $\overline{m}^{t1}_{(l)}$ can be obtained as

$$\gamma_{B_l^{t1}}(\hat{\overline{x}}^{t1(\max)}_{(l)}) = \gamma_{B_l^{t1}}(\overline{m}^{t1}_{(l)}) = 1\,. \quad (39)$$

So, the conditions $\underline{\hat{x}}^{t1(\min)}_{(l)} = \underline{m}^{t1}_{(l)}$ and $\hat{\overline{x}}^{t1(\max)}_{(l)} = \overline{m}^{t1}_{(l)}$ are met for all input variables $\hat{x}^{t1}$, $t_1 = 1,\ldots,m_1$.

Let us consider a technique of learning the consequents of the rules. The variables y_l, $l = 1,\ldots,c$, are the consequents of the fuzzy rules (1), represented by the fuzzy sets C^l_l, $l = 1,\ldots,c$, with their membership functions $\gamma_{C_l^l}(y_l)$. These fuzzy sets C^l_l, $l = 1,\ldots,c$, can be defined on the interval of membership degrees [0,1] and these fuzzy sets can be presented as follows: $C^l_l = (\alpha, \underline{\mu}_l, \overline{\mu}_l, 1)$, where α is the tolerance threshold, $\underline{\mu}_l = \min_{x_i \in A^l_\alpha} \mu_{li}$ and $\overline{\mu}_l = \max_{x_i \in A^l_\alpha} \mu_{li}$. This case is presented in Fig. 2. On the other hand, if $A^l_{(\alpha)}$ and $A^m_{(\alpha)}$, $l \neq m$, are two particularly separated fuzzy α-clusters, then the condition $w \neq 0$ is met in the equation (21). So, a fuzzy set $C^l_m = (0, 1-\overline{\mu}_m, 1-\underline{\mu}_m, 1-\alpha)$ is the consequent for the variable y_m of the l-th fuzzy rule for the case of a low membership degree. The corresponding case is illustrated by Fig. 3.

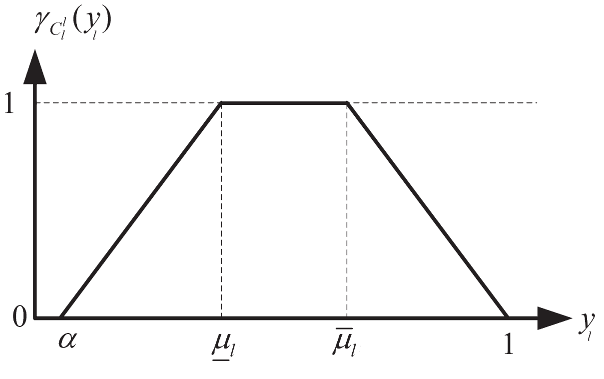

Fig. 2. A membership function for a consequent fuzzy set in a case of a high belonginess

Suppose that the membership functions $\gamma_{C_l^l}(y_l)$ of the fuzzy sets C^l_l, $l = 1,\ldots,c$, are trapezoidal.

The trapezoidal membership functions $\gamma_{C_l^l}(y_l)$ for the fuzzy sets C^l_l, $l = 1,\ldots,c$, can be constructed on the basis of the clustering results. The empty set $A^l_\alpha = \varnothing$, $l \in \{1,\ldots,c\}$, can correspond to some output variable y_l, $l \in \{1,\ldots,c\}$. So, the empty fuzzy set C^l_l will correspond to the output variable y_l, $l \in \{1,\ldots,c\}$, and $\gamma_{C_l^l}(y_l) = 0$ will be the membership function of the corresponding fuzzy set C^l_l.

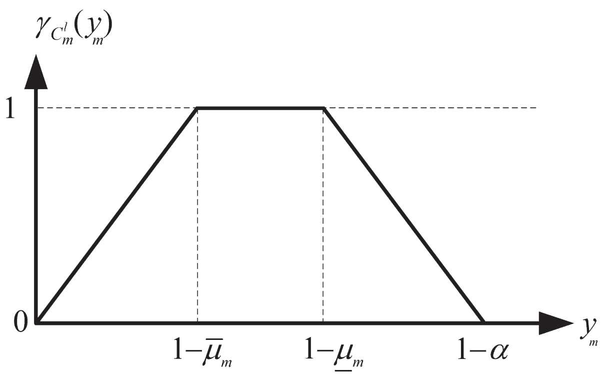

Fig. 3. A membership function for a consequent fuzzy set in the case of a low degree of membership

A scheme of rapid prototyping of the fuzzy inference system from the three-way data can be described shortly as follows: a stationary clustering structure [16] should be constructed in the first step and fuzzy rules must be derived in the second step using the proposed technique.

3. An Illustrative Example

The Sato and Sato [7] three-way data are described in the first subsection. Illustrative examples of data preprocessing are also considered in the first subsection. The second subsection includes results of numerical experiments for three distance functions.

3.1. The Sato and Sato Three-way Data

The Sato and Sato[7] artificial three-way data are a follow-up of a survey on human physical constitution, involving the height, weight, chest girth and sitting height, which are the measurements of 38 boys at three instants, that is, at the age of 13, 14 and 15 years. These data, which originally appeared in Sato and Sato [7], can be rewritten as shown in Table 1.

Table 1. Physical constitution of 38 boys

Boys	Height, cm			Weight, kg			Chest girth, cm			Sitting height, cm		
	13 years old	14 years old	15 years old	13 years old	14 years old	15 years old	13 years old	14 years old	15 years old	13 years old	14 years old	15 years old
1	147	157	162	40	47	54	70	76	81	80	85	87
2	161	166	167	49	50	52	75	75	79	85	87	88
3	153	159	161	45	48	51	72	75	75	86	90	92
4	155	163	168	51	58	66	77	82	87	85	87	92
5	160	165	167	51	56	61	75	77	82	86	88	89
6	153	159	167	38	43	44	67	70	71	81	84	87
7	166	169	172	67	72	79	86	89	92	89	90	95
8	168	174	175	55	60	65	76	79	81	91	93	95
9	142	149	157	35	39	46	69	68	75	75	78	82
10	151	160	165	44	51	57	72	78	80	79	85	89
11	164	167	169	55	58	65	77	79	80	88	89	93
12	153	163	168	42	46	53	70	73	78	83	88	91
13	148	158	164	41	47	51	72	77	81	78	82	85
14	164	169	171	75	84	88	92	97	102	90	93	95
15	145	151	162	34	39	45	65	68	72	76	80	84
16	151	159	162	51	57	64	80	83	87	81	85	87
17	145	153	162	50	55	59	82	84	82	79	81	86
18	154	163	169	47	53	56	71	75	80	82	86	89
19	156	166	171	48	50	56	73	72	75	81	86	89
20	144	149	157	30	33	37	60	62	66	73	75	79
21	154	164	169	41	49	56	69	76	77	82	88	91
22	155	165	169	43	52	57	71	75	79	82	87	90
23	155	162	166	48	58	60	76	85	84	82	86	89
24	155	162	172	49	55	57	73	76	76	80	84	87
25	156	163	164	48	53	54	76	79	82	81	86	87
26	156	164	172	50	53	56	74	76	79	81	84	87
27	162	168	170	45	48	52	71	71	75	84	88	89
28	147	154	163	37	43	50	71	75	80	79	82	86
29	149	157	166	40	47	53	71	79	78	80	83	87
30	148	155	162	37	41	47	69	70	74	78	81	85
31	156	163	166	52	57	62	75	79	81	83	87	89
32	141	151	159	35	42	48	68	74	79	73	77	82
33	140	147	157	30	34	43	67	70	73	76	77	83
34	146	153	161	49	52	53	76	78	76	80	79	84
35	162	168	161	53	58	53	74	78	76	86	79	84
36	146	158	165	36	44	51	68	75	73	77	85	89
37	141	151	158	41	46	51	71	75	76	76	80	83
38	158	167	171	65	71	79	93	93	90	85	90	91

Denote the height by $\hat{x}^1$, the weight by $\hat{x}^2$, the chest girth by $\hat{x}^3$ and the sitting height by $\hat{x}^4$. Each attribute $\hat{x}^{t1}$, $t_1 = 1,\ldots,4$, is observed at three instants (ages), $t_2 = 1,\ldots,3$. The value of the t_1-th attribute at the t_2-th moment for the i-th object is denoted by $\hat{x}_i^{t1(t2)}$, $i = 1,\ldots,38$, $t_1 = 1,\ldots,4$, $t_2 = 1,\ldots,3$. The data were preprocessed according to formulae (26), (27) and (29).

3.2. Experimental Results

Let us consider results of the application of the proposed technique to the Sato and Sato artificial three-way data. The data was processed by the D-AFC(c)-algorithm with the number of fuzzy clusters $c = 2,3\ldots$, using the measure of separation and compactness of an allotment (24). The performance of the validity measure is shown in Fig. 4.

The optimal number of fuzzy clusters is equal 3 and this number corresponds to the first minimum of the measure of separation and compactness of the allotment. The corresponding allotment $R^*(X)$ among three fully separate fuzzy clusters was obtained for the tolerance threshold $\alpha = 0.93120$. The membership functions of three classes of the allotment are presented in Fig. 5 and he values which equal zero are not shown in this figure. The membership values of the first class are represented by +, the membership values of the second class are represented by ■, and the membership values of the third class are represented by ×.

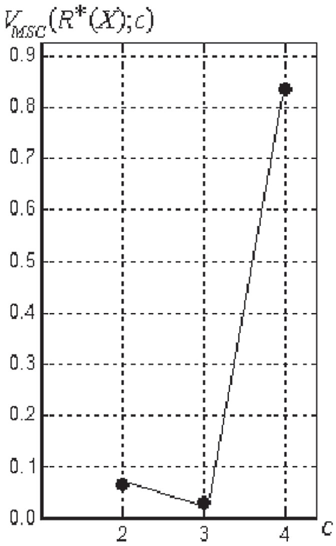

Fig. 4. Plot of the measure of separation and compactness for the Sato and Sato three-way data set

So, by executing the D-AFC(c)-algorithm for three classes, we obtain that the first class is formed by 3 elements, the second class by 7 elements, and the third class by 28 elements. The value of the membership function of the fuzzy cluster, which corresponds to the first class, is maximal for the fourteenth object and is equal 0.98298. So, the fourteenth object is a typical point of the fuzzy cluster which corresponds to the first class. The membership value of the twentieth object is equal 0.97888 and this value is maximal for the fuzzy cluster which corresponds to the second class. Thus, the twentieth object is a typical point of the fuzzy cluster which corresponds to the second class. The membership function of the third fuzzy cluster is maximal for the fifth object and is equal 0.98392. That is why the fifth object is a typical point of the fuzzy cluster which corresponds to the third class.

We could see that the boys in the first cluster have a good physical constitution through all three years of age. Conversely, the boys in the second cluster have a poor constitution. The boys who belong to the third cluster have a standard constitution. So, the results, which are obtained from the D-AFC(c)-algorithm are similar to the results, which were obtained by Sato and Sato [7] using their multicriteria optimization method.

The rule base induced by the proposed technique from the clustering result obtained by using the D-AFC(c)-algorithm can be seen in Fig. 6 – Fig. 8. In particular, the performance of the fuzzy inference system for the thirty-second boy at the first time measurement is presented in Fig. 6.

The total area is zero while using the defuzzification procedure for the output variables Class 1 and

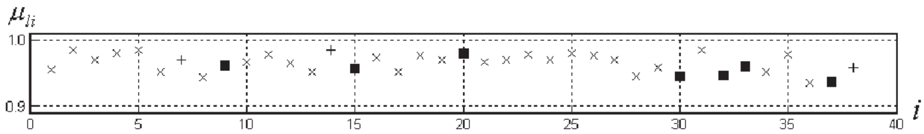

Fig. 5. Membership functions of three fuzzy clusters obtained from the D-AFC(c)-algorithm

Table 2. Results of performance of the generated fuzzy inference system for the data set

Classes	Times measurement		
	13 years old	14 years old	15 years old
1	7, 14, 38	7, 14, 38	4, 7, 14, 38
2	1, 2, 6, 9, 10, 12, 13, 15, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 33, 34, 36	1, 6, 9, 15, 20, 28, 30, 32, 33, 36, 37	9, 15, 20, 30, 33, 37
3	1, 2, 3, 4, 5, 6, 8, 10, 11, 12, 13, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 34, 35, 36, 37	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37

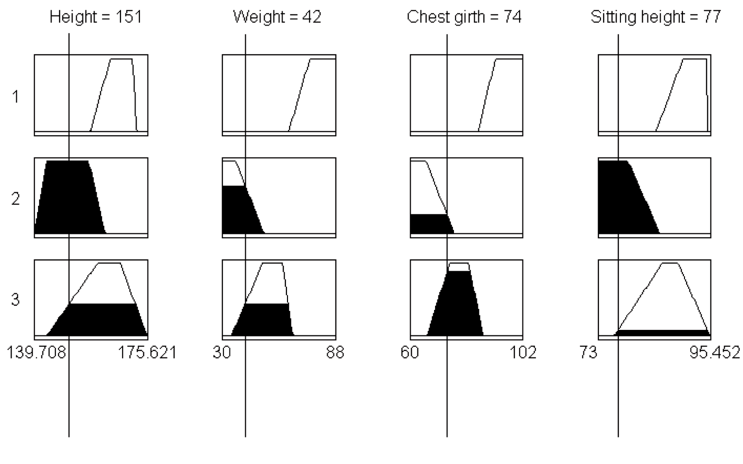

Fig. 6. The performance of the generated fuzzy inference system for the thirty-second boy at the first time measurement (13 years old)

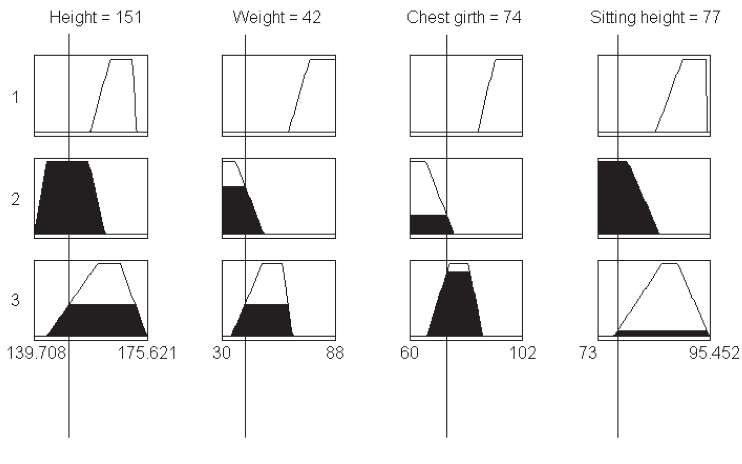

Fig. 7. The graph of performance of the generated fuzzy inference system for the thirty-second boy at the second time measurement (14 years old)

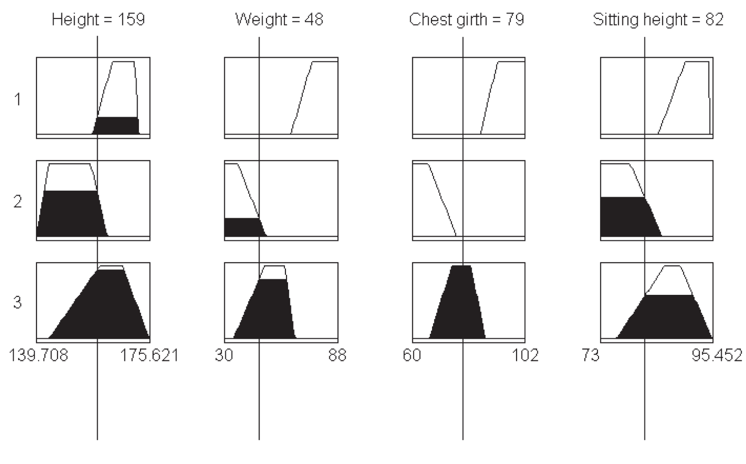

Fig. 8. The graph of performance of the generated fuzzy inference system for the thirty-second boy at the third time measurement (15 years old)

Class 3. That is why the average values of the range of the output variables Class 1 and Class 3 are used as the output values and these values are equal 0.5. These values can be interpreted as uncertain membership degrees.

The performance of the fuzzy inference system for the thirty-second boy at the second time measurement is presented in Fig. 7. It should be noted, that at that time the boy belonged to the second and the third classes.

The performance of the fuzzy inference system for the thirty-second boy at the third time measurement is presented in Fig. 8.

So, we could see that the thirty-second boy has a tendency of growth during the period from 14 years old to 15 years old.

The results of the numerical experiment for all 38 boys at all three times measurement are summarized in Table 2.

The Sato and Sato three-way data were classified using the fuzzy inference system constructed. Evidently, the results obtained are correlated with the results obtained from the D-AFC(c)-algorithm. So, the fuzzy inference system is accurate. On the other hand, we can observe the trend of development of each boy.

The result which is obtained from the fuzzy inference system can easily be interpreted. Thus, the obtained model is suitable for the interpretation since the consequents of the rules are the same or close to the current class labels, such that each rule can be taken to describe all classes.

4. Conclusions

Many techniques to design fuzzy inference systems from data are available; basically, they all take advantage, explicitly or implicitly, of the property of the fuzzy inference systems to be the universal approximators. This paper presents an extension of an automatic method to design fuzzy inference system for classification via heuristic possibilistic clustering. This method can be considered as an approach to rapid prototyping of the fuzzy inference systems for the case of the three-way data. The proposed method is simple in comparison with other well-known approaches. The results obtained for the well=known Sato and Sato three-way data set show the effectiveness of the proposed method.

Some approaches, such as those based on the use of genetic algorithms or neuro-fuzzy techniques can be used for fuzzy rules tuning. On the other hand, a scheme of on-line training of the fuzzy inference system can be developed. These perspectives for further research are of a great interest both from the theoretical and practical points of view.

Acknowledgements

The research has been partially supported, first, by the National Centre of \science under Grant No. UMO-2012/05/B/ST6/03068 (J. Kacprzyk, J. W. Owsiński), and secod, by the Mianowski Fund and the Ministry of Foreign Affairs of the Republic of Poland (D. Viattchenin). The authors are grateful to Mr. Aliaksandr Damaratski for developing software to implement ad test the tools developed.

AUTHORS

Janusz Kacprzyk, Jan W. Owsinski – Systems Research Institute, Polish Academy of Sciences, 6 Newelska St., 01-447 Warsaw, Poland.
E-mails: {kacprzyk, owsinski}@ibspan.waw.pl

Dmitri A. Viattchenin* – United Institute of Informatics Problems, National Academy of Sciences of Belarus, 6 Surganov St., 220012 Minsk, Belarus.
E-mail: viattchenin@mail.ru

*Corresponding author

References

[1] Bezdek J.C., *Pattern Recognition with Fuzzy Objective Function Algorithms,* Plenum Press, New York, 1981. DOI: http://dx.doi.org/10.1007/978-1-4757-0450-1
[2] Höppner F., Klawonn F., Kruse R., Runkler T., *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*, Wiley, Chichester, 1999.
[3] Kacprzyk J., *Multistage Fuzzy Control*, Wiley, Chichester, 1997.
[4] Krishnapuram R., Keller J.M., "A possibilistic approach to clustering", *IEEE Trans. on Fuzzy Syst.*, no. 1, 1993, 98–110. DOI: http://dx.doi.org/10.1109/91.227387
[5] Mamdani E.H., Assilian S., "An experiment in linguistic synthesis with a fuzzy logic controller", *International Journal of Man-Machine Studies*, vol. 7, no. 1, 1975, 1–13. DOI: http://dx.doi.org/10.1016/S0020-7373(75)80002-2
[6] Mandel I.D., *Clustering Analysis*, Finansy i Statistica, Moscow, 1988. (in Russian)
[7] Sato M., Sato Y., "On a multicriteria fuzzy clustering method for 3-way data", *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 2, no. 2, 1994, 127–142. DOI: http://dx.doi.org/10.1142/S0218488594000122
[8] Viattchenin D.A., "A new heuristic algorithm of fuzzy clustering", *Control & Cybernetics*, vol. 33, 2004, 323–340.
[9] Viattchenin D.A., "A direct algorithm of possibilistic clustering with partial supervision", *Journal of Automation, Mobile Robotics and Intelligent Systems*, vol. 1, no. 3, 2007, 29–38.
[10] Viattchenin D.A., "Direct algorithms of fuzzy clustering based on the transitive closure operation and their application to outliers detection", *Artificial Intelligence*, vol. 3, 2007, 205–216. (in Russian)
[11] Viattchenin D.A., "An outline for a heuristic approach to possibilistic clustering of the three-way data", *Journal of Uncertain Systems*, vol. 3, 2009, 64–80.
[12] Viattchenin D.A., "An algorithm for detecting the principal allotment among fuzzy clusters and its application as a technique of reduction of analyzed features space dimensionality", *Journal of Information and Organizational Sciences*, vol. 33, no. 1, 2009, 205–217.
[13] Viattchenin D.A., Automatic generation of fuzzy inference systems using heuristic possibilistic clustering, *Journal of Automation, Mobile Robotics and Intelligent Systems*, vol. 4, no. 3, 2010, 36–44.
[14] Viattchenin D.A., "Derivation of fuzzy rules from interval-valued data", *International Journal of Computer Applications*, vol. 7, no. 3, 2010, 13–20. DOI: http://dx.doi.org/10.5120/1146-1500
[15] Viattchenin D.A., "Validity measures for heuristic possibilistic clustering", *Information Technology and Control*, vol. 39, 2010, 321–332.
[16] Viattchenin D.A., "Constructing stable clustering structure for uncertain data set", *Acta Electrotechnica et Informatica*, vol. 11, no. 3, 2011, 42–50. DOI: http://dx.doi.org/10.2478/v10198-011-0028-5.