A PROXIMITY-BASED FUZZY CLUSTERING FOR FUZZY NUMBERS

D.A. Viattchenin
National Academy of Sciences of Belarus, United Institute of Informatics Problems
Minsk, Belarus
e-mail: viattchenin@mail.ru

Abstract

The paper deals in a preliminary way with the problem of fuzzy clustering of fuzzy data. A distance between triangular fuzzy numbers is considered and the data preprocessing method-ology for constructing of a fuzzy tolerance matrix is described. A numerical example is given and results of application of D-$AFC(c)$-algorithm of fuzzy clustering to a set of triangular fuzzy numbers are considered in the example. Some preliminary conclusions are stated.

1 Introduction

Most fuzzy clustering techniques are designed for handling crisp data with their class membership functions. However, the data can be uncertain or fuzzy. Fuzzy numbers are well used to model the fuzziness of data. Yang and Ko [2] recently proposed a class of fuzzy c-number (FCN) clustering procedures for fuzzy data clustering. A direct fuzzy clustering method was outlined by Viattchenin [1], where a basic version of direct fuzzy clustering algorithm was described. Detection of a unique allotment among given number c of fuzzy α-clusters is the aim of classification. The version of the algorithm, which is described by Viattchenin [1], can be called the D-$AFC(c)$-algorithm.

The goal of the paper is a consideration of possibilities of an application of object-data clustering techniques to fuzzy data clustering. For this purpose, a consideration of distance between triangular fuzzy numbers is presented. A methodology for the data preprocessing is described. Results of application of the D-$AFC(c)$-algorithm to a set of triangular fuzzy numbers are considered in a numerical example. Preliminary conclusions are formulated.

2 A distance between triangular fuzzy numbers

Triangular fuzzy numbers are important case of LR-type fuzzy numbers. The problem of proximity-based fuzzy clustering for fuzzy data can be illustrated on a simple example of triangular fuzzy numbers. That is why a concept of a LR-type fuzzy number must be defined in the first place.

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

$$\mu_V(x) = \begin{cases} L\left(\dfrac{m-x}{a}\right), & for \;\; x \le m \\ R\left(\dfrac{x-m}{b}\right), & for \;\; x \ge m \end{cases}, \tag{1}$$

where m is called the mean value of V and a and b are called the left and right spreads.

In LR-type fuzzy numbers, the triangular and Gaussian fuzzy numbers are most commonly used. In particular, for a LR-type fuzzy number $V = (m, a, b)_{LR}$ if L and R are of the form

$$T(x) = \begin{cases} 1-x, & 0 \le x \le 1 \\ 0, & otherwise \end{cases}, \tag{2}$$

then V is called a triangular fuzzy number, denoted by $V = (m, a, b)_T$ and its membership function is defined as

$$\mu_V(x) = \begin{cases} 1 - \dfrac{m-x}{a}, & for \;\; x \le m, \; (a > 0) \\ 1 - \dfrac{x-m}{b}, & for \;\; x \ge m, \; (b > 0) \end{cases}. \tag{3}$$

Let us consider a distance between triangular fuzzy numbers. The distance was proposed by Yang and Ko [2]. Let $\mathcal{F}_{(LR)FN}(\Re)$ denote the set of all LR-type fuzzy numbers and $X = V_1, \ldots, V_n$ be a set of fuzzy numbers in $\mathcal{F}_{(T)FN}(\Re)$. A distance $d^2_{(T)FN}(V_i, V_j)$ for any two triangular fuzzy numbers $V_i = (m_i, a_i, b_i)_T$ and $V_j = (m_j, a_j, b_j)_T$ in the space $\mathcal{F}_{(T)FN}(\Re)$ can be defined as follows:

$$\begin{aligned} d^2_{(T)FN}(V_i, V_j) = (m_i - m_j)^2 + \left((m_i - m_j) - \frac{1}{2}(a_i - a_j)\right)^2 + \\ + \left((m_i - m_j) + \frac{1}{2}(b_i - b_j)\right)^2. \end{aligned} \tag{4}$$

After application of a distance to the data set $X = \{V_1, \ldots, V_n\}$ a matrix of coefficients of pair wise dissimilarity between objects $d_{n \times n} = [d_{ij}]$, $i, j = 1, \ldots, n$ can be obtained.

3 The data preprocessing

Some object-data fuzzy clustering procedures can be applied directly to the data given as the matrix of dissimilarity coefficients $d_{n \times n} = [d_{ij}]$, $i, j = 1, \ldots, n$. A proximity relation is known as a tolerance relation also. Fuzzy tolerance is the fuzzy binary intransitive relation which satisfies a symmetricity property and a reflexivity property. A matrix of fuzzy intolerance $I = [\mu_I(x_i, x_j)]$, $i, j = 1, \ldots, n$ must be obtained for construction of the fuzzy tolerance matrix. For the purpose, the data can be normalized as follows:

$$\mu_I(x_i, x_j) = \frac{d_{ij}}{\max\limits_{i,j} d_{ij}}, \tag{5}$$

where d_{ij}, $i, j = 1, \ldots, n$ are dissimilarity coefficients. The matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, n$ can be obtained after application of complement operation

$$\mu_T(x_i, x_j) = 1 - \mu_I(x_i, x_j), \ \forall i, j = 1, \ldots, n \tag{6}$$

to the matrix of fuzzy intolerance $I = [\mu_I(x_i, x_j)]$, $i, j = 1, \ldots, n$. The matrix of fuzzy tolerance $T = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, n$ is the matrix of initial data for the D-$AFC(c)$-algorithm and some other proximity-based fuzzy clustering techniques.

4 A numerical example

Let us consider an application of the D-$AFC(c)$-algorithm to the classification problem for the following illustrative example, which was considered by Yang and Ko [2]. The set $X = \{V_1, \ldots, V_{30}\}$ of triangular fuzzy numbers is presented in Figure 1.

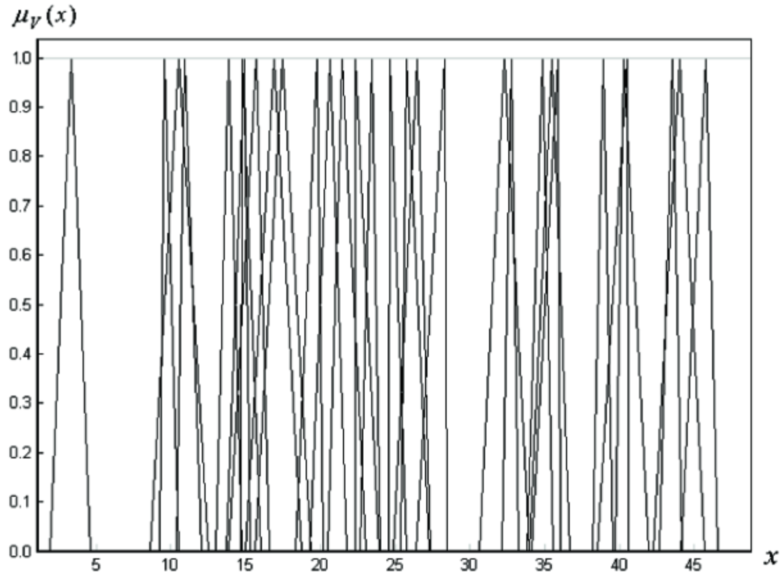

Figure 1: The graph of 30 triangular fuzzy numbers

The distance $d^2_{(T)FN}(V_i, V_j)$ was applied to the data set $X = \{V_1, \ldots, V_{30}\}$ and the matrix $T = [\mu_T(x_i, x_j)]$, $i, j = 1, ..., 30$ of fuzzy tolerance was obtained. By executing the D-$AFC(c)$-algorithm for three classes we obtain following: the first class is formed by 4 elements; the second class by 17 elements; the third class by 9 elements. The allotment, which corresponds to the result, was obtained for the tolerance threshold $\alpha = 0.93455116$. The fuzzy number $V_1 = (m_1 = 3.34,\ a_1 = 1.46, b_1 = 1.30)_T$ is a typical point τ^1 of the fuzzy α-cluster which corresponds to the first class. The fuzzy number $V_{15} = (m_{15} = 23.47, a_{15} = 0.81, b_{15} = 0.51)_T$ is the typical point τ^{15} of the fuzzy α-cluster which corresponds to the second class and the fuzzy number $V_{30} = (m_{30} = 45.77, a_{30} = 1.71, b_{30} = 0.79)_T$ is the typical point τ^{30} of the fuzzy α-cluster which corresponds to the third class. Member-ship functions of three classes are shown in Figure 2.

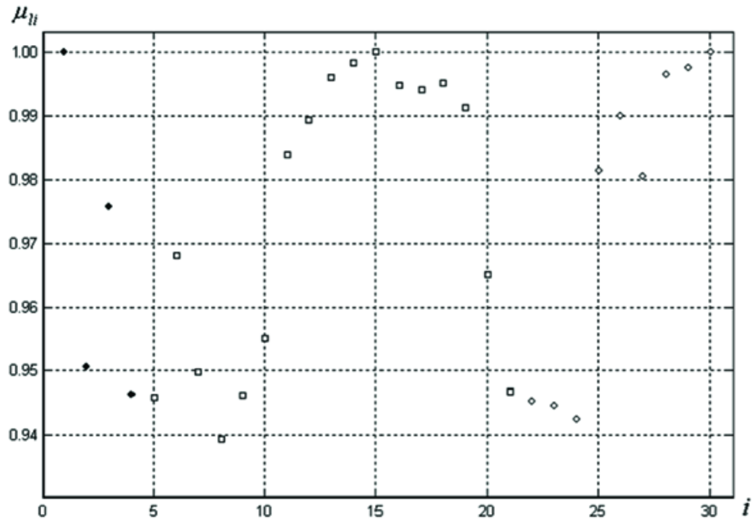

Figure 2: Membership functions of three classes of the allotment

5 Concluding remarks

The result of the numerical experiment seems to be satisfactory. The result of application of the $D\text{-}AFC(c)$-algorithm to the set of triangular fuzzy numbers shows that the $D\text{-}AFC(c)$-algorithm is an effective clustering procedure for classification of fuzzy numbers.

Acknowledgements

I would like to thank to Dr. Jan W. Owsinski and Prof. Janusz Kacprzyk for their useful remarks during the paper preparation. I am grateful to the Mianowski Fund and Siemens for financial support of the investigation in the Systems Research Institute of the Polish Academy of Sciences. I would like to thank to Mr. A. Damaratski, Mr. D. Novikau and Mr. E. Snezhko for elaboration the experimental software.

References

[1] Viattchenin D.A. (2004). A new heuristic algorithm of fuzzy clustering. *Control & Cybernetics*. Vol. **33**, pp. 323-340.

[2] Yang M.-S., Ko C.-H. (1996). On a class of fuzzy c-numbers clustering procedures for fuzzy data. *Fuzzy Sets and Systems*. Vol. **84**, pp. 49-60.