Journal of Mathematical Modeling
Vol. 6, No. 2, 2018, pp. 123-136

JMM

A new outlier detection method for high dimensional fuzzy databases based on LOF

Alireza Fakharzadeh Jahromi[†][*]**and Zahra Ebrahimi Mimand**[‡]

[†]*Department of Applied Mathematics, Shiraz University of Technology*
Shiraz, Iran
[†]*Fars Elites Foundation, Shiraz, Iran, P.O. Box 71966-98893*
[‡]*Department of Mathematics, Faculty of Basic Science,*
PayamNoor University, Shiraz Branch, shiraz, Iran
emails: A_fakharzadeh@sutech.ac.ir, zahra_ebrahimi26@yahoo.com

Abstract. Despite the importance of fuzzy data and existence of many powerful methods for determining crisp outliers, there are few approaches for identifying outliers in fuzzy database. In this regard, the present article introduces a new method for discovering outliers among a set of multidimensional data. In order to provide a complete fuzzy strategy, first we extend the density-based local outlier factor method (LOF), which is successfully applied for identifying multidimensional crisp outliers. Next, by using the left and right scoring defuzzyfied method, a fuzzy data outlier degree is determined. Finally, the efficiency of the method in outlier detection is shown by numerical examples.

Keywords: Fuzzy numbers, Outlier data, LOF factor, α-cut, Left and right scoring.
AMS Subject Classification: 34A34, 65L05.

1 Introduction and background

Identifying outlier data is a considerable and remarkable concept in discovering rare events, fraud, diversion of the majority or identifying

[*]Corresponding author.

Received: 23 July 2017 / Revised: 2 March 2018 / Accepted: 10 June 2018.
DOI: 10.22124/jmm.2018.8102.1108

special cases; for example, a typical business application such as: analyzing market management, analyzing market basket, objective marketing, understanding costumer's behavior, risk analysis and management, fraud discovery including: telephone frauds, insurance fraud, credit cards fraud, fraud in bank data, data mining: such as text clarification (emails, news groups, etc.), medical such as: discovering malignant tumors in data set relating to benign and malignant tumors, DNA arrangement analysis, medical images, web mining such as proposing related pages, improving search machines, discovering crime in electronic industry, discovering suspicious accounts in accounting and leveling data related to designs and image processing [2,4,5,8].

In spite of the large variety of methods for outlier detection in Crisp data and the importance of fuzziness real world, there are few methods on discovering fuzzy outliers among a set of fuzzy high dimensional data. Since among existing approaches, LOF is one of the most effective methods in detecting high dimensional data ([2,4]) providing an extension of LOF method for identifying outliers in a fuzzy database, would be the aim of this paper.

It is necessary to remind that most early studies in the field of outlier discovering data are done in the statistics domain. These studies can be classified in two categories: distribution and depth. In the first category, outlier data detection methods have been based on possible unknown main distribution [3]. The second category is not effective for high dimensional data, because it relies on convex shell space distribution (that has a lower bound distribution); Knorr and Ng suggested a discovering outlier data concept, based on the space [9]. According to them, using distribution basic methods have had better result, and of course more complexity, in comparison with the depth basic methods. Even in [9] discovering outliers by using the concept of K-neighborhood is developed, but it also was based on spaces as well.

LOF is an outlier detection method in which, as being introduced by [2], it became the basic method in identifying outlier data based on density; indeed, a newer and more developed versions of it were provided for LOF in 2003. For instance, LOF$'$ and LOF$''$ techniques were provided two new definitions of outlier factor, and also GridLOF method was an algorithm, that adds one step to the previous LOF algorithm [4]. It should be noted that although other density based methods such as LOCI (Local Correlation Integral) [10] were introduced after LOF, but still the known LOF method is the most functional density-based methods for identifying outliers. Regarding the aim of this paper, it is necessary to remind that even recently, some attempts for using fuzzy methods have been done, (see, for instance [7]) but they never used the LOF method. By considering the powerful ability of LOF outlier detection method, in this paper we are going to extend this useful method, so that it is able to identify the outliers in a multi-dimensional fuzzy data set. The rest of the paper is organized as follow. In Section 2, we review some basic fuzzy concepts, which are very useful in our discussions. In Section 3, by extending the LOF method, the new algorithm is presented. To show the efficiency of our algorithm, some numerical results are presented in Section 4. Finally, some concluding remarks are given in Section 5.

2 Preliminary and basic concepts

In this section first some basic concepts and definitions are given; then the right and left scoring method (using for fuzzy elimination) is introduced. Finally, we discuss about a way to display fuzzy numbers by using α-cuts. More details on the topics of this section can be found in [2,6,9,11,13,14].

2.1 Some definitions

To explain the new method for discovering outliers in fuzzy data set, it is necessary to have a common language and concepts. In this subsection we present the prerequisite definitions and concepts from [2]. Also, we explain some important fuzzy concepts that we will use in rest of this paper.

Definition 1. (k-distance): For any positive integer k, the k-distance of object p, denoted by k-distance (p), is the distance $d(p, o)$ between p and an object $o \in D$ such that:

(i) for at least k objects $o' \in D \setminus \{p\}$ we have $d(p, o') \leq d(p, o)$;

(ii) for at most $k - 1$ objects $o' \in D \setminus \{p\}$ we have $d(p, o') < d(p, o)$.

Definition 2. (k-distance neighborhood): The k-distance neighborhood of p contains every object whose distance from p is not greater than the k-distance, i.e.

$$Nk - distance(p) = \big\{q \in D \setminus \{p\} |\ d(p, q) \leq k - distance(p)\big\};$$

These objects q are called the k-nearest neighbors of p.

Definition 3. *(reachability distance): For $k \in N$ the reachability distance of object p with respect to object o is defined as:*

$$reach - distk(p, o) = \max\big\{k - distance(o), d(p, o)\big\}.$$

Definition 4. *(local reachability density): The local reachability density of p is defined as*

$$lrd_{minpts}(p) = \frac{|N_{minpts}(p)|}{\sum_{o \in N_{minpts}(p)} reach - dist_{minpts}(p, o)}.$$

Definition 5. *(outlier factor): The (local) outlier factor of p is defined as*

$$LoF_{minpts}(p) = \frac{\sum_{o \in N_{minpts}(p)} \frac{lrd_{minpts}(o)}{lrd_{minpts}(p)}}{|N_{minpts}(p)|}.$$

To illustrate the fuzzy concepts and definitions, it is supposed that the readers are familiar with fuzzy set, fuzzy numbers and membership functions (for more detail please see [14]).

Definition 6. *For $0 \leq \alpha \leq 1$, an α-cut of a fuzzy set $\tilde{A}$, A_α is defined as $A_\alpha = \{x : \mu_{\tilde{A}}(x) \geq \alpha\}$, where $\mu_{\tilde{A}}$ is the membership function of $\tilde{A}$; since $\mu_{\tilde{A}}$ is bounded, an α-cut actually is an interval on the x-axis, like $A_\alpha = [a_1^{(\alpha)}, a_2^{(\alpha)}]$.*

Definition 7. Addition and subtraction operations of two fuzzy numbers $\tilde{A}$ and $\tilde{B}$ in terms of α- cuts are defined as follows:

$$\begin{aligned}(\tilde{A})_\alpha + (\tilde{B})_\alpha &= \left[a_1^{(\alpha)}, a_2^{(\alpha)}\right] + \left[b_1^{(\alpha)}, b_2^{(\alpha)}\right] = \left[a_1^{(\alpha)} + b_1^{(\alpha)}, a_2^{(\alpha)} + b_2^{(\alpha)}\right];\\ (\tilde{A})_\alpha - (\tilde{B})_\alpha &= \left[a_1^{(\alpha)}, a_2^{(\alpha)}\right] - \left[b_1^{(\alpha)}, b_2^{(\alpha)}\right] = \left[a_1^{(\alpha)} - b_1^{(\alpha)}, a_2^{(\alpha)} - b_2^{(\alpha)}\right].\end{aligned}$$

Definition 8. The minimum of numbers $\tilde{A}$ and $\tilde{B}$, $\tilde{A} \wedge \tilde{B}$. and the maximum of them, $\tilde{A} \vee \tilde{B}$, are defined in terms of α-cuts as follows:

$$\begin{aligned}(\tilde{A})_\alpha \vee (\tilde{B})_\alpha &= \left[a_1^{(\alpha)}, a_2^{(\alpha)}\right] \vee \left[b_1^{(\alpha)}, b_2^{(\alpha)}\right] = [\min\{a_1^{(\alpha)}, b_1^{(\alpha)}\}, \min\{a_2^{(\alpha)}, b_2^{(\alpha)}\}];\\ (\tilde{A})_\alpha \wedge (\tilde{B})_\alpha &= \left[a_1^{(\alpha)}, a_2^{(\alpha)}\right] \wedge \left[b_1^{(\alpha)}, b_2^{(\alpha)}\right] = \left[\max\{a_1^{(\alpha)}, b_1^{(\alpha)}\}, \max\{a_2^{(\alpha)}, b_2^{(\alpha)}\}\right].\end{aligned}$$

Definition 9. A triangular fuzzy number like $\tilde{M}(l, m, u)$ is illustrated with the following membership function:

$$\mu_{\tilde{M}}(x) = \begin{cases} \frac{x-l}{m-l}, & l < x < m, \\ \frac{(u-x)}{(u-m)}, & m < x < u, \\ 0, & otherwise, \end{cases}$$

where l and u are the lower and upper bounds for $\tilde{M}$, respectively.

Also some algebraic operations between triangular numbers $\tilde{M} = (l.m.u)$ and $\tilde{N} = (a, b, c)$, are introduced as follow: Reverse of the triangular fuzzy number M: $\tilde{M}^{-1} = (\frac{1}{u}, \frac{1}{m}, \frac{1}{l})$.

Multiplication of the two triangular fuzzy numbers:

$$\begin{aligned} if\ \tilde{M} < 0,\ \tilde{N} < 0,\ \text{then}\ \ \tilde{M} \times \tilde{N} = (uc, mb, la); \\ if\ \tilde{M} < 0,\ \tilde{N} > 0,\ \text{then}\ \ \tilde{M} \times \tilde{N} = (lc, mb, ua); \\ if\ \tilde{M} > 0,\ \tilde{N} > 0,\ \text{then}\ \ \tilde{M} \times \tilde{N} = (la, mb, uc). \end{aligned}$$

Definition 10. Suppose $\tilde{A}_i = (a_1^i, a_2^i, a_3^i)$ $(i = 1, 2, \ldots, n)$ is a set of triangular fuzzy numbers; a weighted average for $\tilde{A}_i$'s could be defined as:

$$\tilde{A}_{Ave} = \left(\frac{1}{n}\sum_{i=1}^{n} a_1^i, \frac{1}{n}\sum_{i=1}^{n} a_2^i, \frac{1}{n}\sum_{i=1}^{n} a_3^i\right).$$

Also another weighted average of triangular fuzzy numbers for weights W_i's in $[0, 1]$, where $\sum_{i=1}^{n} W_i = 1$, is introduced as follow [11, 12]:

$$\tilde{A}_{Ave}^{W} = \big(\sum_{i=1}^{n} W_i a_1^i, \sum_{i=1}^{n} W_i a_2^i, \sum_{i=1}^{n} W_i a_3^i\big) \ \ (i = 1, 2, \ldots, n).$$

Note: Let M be a set containing the triangular fuzzy numbers $\tilde{A}_i$ $(i = 1, 2, \ldots, n)$, such that any $\tilde{A}_i$ is an m-dimensional fuzzy vector where its components are symmetric triangular fuzzy numbers $\tilde{a}_{ij}$ $(j = 1, 2, \ldots, n)$; then we can write:

$$\begin{aligned} M \ &= \begin{bmatrix} \tilde{A}_1 \\ \tilde{A}_2 \\ \vdots \\ \tilde{A}_n \end{bmatrix} = \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \ldots & \tilde{a}_{1m} \\ \tilde{a}_{21} & \tilde{a}_{22} & \ldots & \tilde{a}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{a}_{n1} & \tilde{a}_{n2} & \ldots & \tilde{a}_{nm} \end{bmatrix} \\ &= \begin{bmatrix} (a_1^{11}, a_2^{11}, a_3^{11}) & (a_1^{12}, a_2^{12}, a_3^{12}) & \ldots & (a_1^{1m}, a_2^{1m}, a_3^{1m}) \\ (a_1^{21}, a_2^{21}, a_3^{11}) & (a_1^{22}, a_2^{22}, a_3^{22}) & \ldots & (a_1^{2m}, a_2^{2m}, a_3^{2m}) \\ \vdots & \vdots & \vdots & \\ (a_1^{n1}, a_2^{n1}, a_3^{n1}) & (a_1^{n2}, a_2^{n2}, a_3^{n2}) & \ldots & (a_1^{nm}, a_2^{nm}, a_3^{nm}) \end{bmatrix}. \end{aligned}$$

Therefore, we have:

$$\begin{bmatrix} (\tilde{A}_1)_{Ave} \\ (\tilde{A}_2)_{Ave} \\ \vdots \\ (\tilde{A}_n)_{Ave} \end{bmatrix} = \begin{bmatrix} (\sum_j W_j^1 a_1^{1j}, \sum_j W_j^1 a_2^{1j}, \sum_j W_j^1 a_3^{1j}) \\ (\sum_j W_j^2 a_1^{2j}, \sum_j W_j^2 a_2^{2j}, \sum_j W_j^2 a_3^{2j}) \\ \vdots \\ (\sum_j W_j^n a_1^{nj}, \sum_j W_j^n a_2^{nj}, \sum_j W_j^n a_3^{nj}) \end{bmatrix},$$

where each $W^i = (W_1^i, W_2^i, \ldots, W_n^i)^T$, $i = 1, 2, \ldots, n$, is a weight vector such that $W_j^i \in [0, 1]$ and $\sum_{j=1}^{n} W_j^i = 1$.

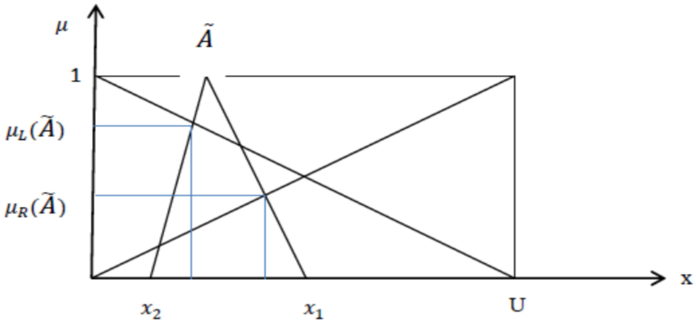

Figure 1: A graphical determination values of the left and right points.

2.2 Fuzzy removal left and right scoring method

In order to transfer a fuzzy number into a crisp one (defuzzyfied), several methods are introduced, for instance center of gravity maximum membership function, or fuzzy number right and left scoring. Due to the use of continuous membership function, in this study, the right and left scoring method is applied for fuzzy elimination.

To explain this method, assume that $\tilde{A} = (\alpha, m, \beta)$ is a triangular fuzzy number; Figure 1 shows the necessary values for the left and right scoring of $\tilde{A}$ graphically [1]:

For the biggest bound of number U, in this figure, x_1 is a cross-section of $\mu_{\max} = \frac{1}{U}x$ in the right half and x_2 is cross-section of $\mu_{min} = 1 - \frac{1}{U}x$ in the left half; in this regard $\mu_L(\tilde{A})$ is called the left score, $\mu_R(\tilde{A})$ is called the right score and $\mu_T(\tilde{A})$, is called the total score which are defined as follows:

$$\mu_R(\tilde{A}) = \frac{m+\beta}{U+\beta}, \; \mu_L(\tilde{A}) = 1 - \frac{m}{U+\alpha}, \; \mu_T(\tilde{A}) = \frac{\mu_R(\tilde{A}) + 1 - \mu_L(\tilde{A})}{2}.$$

In this regard, one may use the amount of $\mu_T(\tilde{A})$ as the defuzziness score of $\tilde{A}$.

2.3 α-Cut performance of a fuzzy number

Before introducing the extended version of LOF algorithm for identifying fuzzy outlier data, we need to discuss about a way to display fuzzy numbers by using α-cuts. We prefer to start this discussion by providing an example; Figure 2 displays a symmetric triangular fuzzy number which is approximately 3. Hence, a membership function for this number is as follows:

$$\mu_{\tilde{3}}(x) = \begin{cases} \frac{1}{2}x - \frac{1}{2}, & 1 \leq x \leq 3, \\ -\frac{1}{2}x + \frac{5}{2}, & 3 \leq x \leq 5, \\ 0 & \mathit{otherwise}. \end{cases}$$

On the other hand, the membership function can be shown by α-cuts (intervals $[a_1^{(\alpha)}, a_2^{(\alpha)}]$) in a parametric representation, as shown in Figure 3, where $0 < \alpha < 1$ is a parameter. In this manner, when $\alpha_1 < \alpha_2$, we have $[a_1^{(\alpha_2)}, a_2^{(\alpha_2)}] \subset [a_1^{(\alpha_1)}, a_2^{(\alpha_1)}]$; in fact, we have different intervals for different amounts of α (see Figure 3). In this order, the mentioned intervals for $\tilde{3}$ can be determined for different values of α as follows.

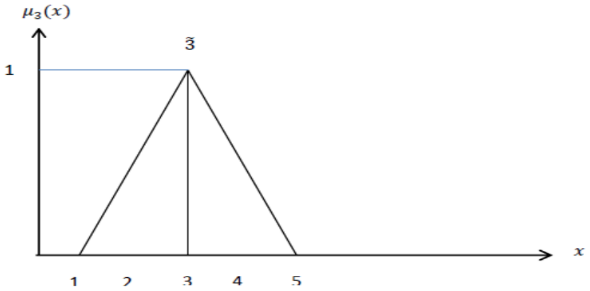

Figure 2: Presentation of the symmetric triangular fuzzy number $\tilde{3}$.

$$\begin{aligned} &\alpha = 0 : [1, 5];\ \alpha = 0.1 : [1.2, 4.8];\ \alpha = 0.2 : [1.4, 4.6];\ \alpha = 0.3 : [1.6, 4.4]; \\ &\alpha = 0.4 : [1.8, 4.2];\ \alpha = 0.5 : [2, 4];\ \alpha = 0.6 : [2.2, 3.8]; \\ &\alpha = 0.7 : [2.4, 3.6];\ \alpha = 0.8 : [2.6, 3.4];\ \alpha = 0.9 : [2.8, 3.2];\ \alpha = 1 : [3, 3]. \end{aligned}$$

Therefore, in discretization scheme, we can display the approximated fuzzy number $\tilde{3}$ by the following set:

$$\left\{ \begin{array}{c} (1.2, 0.1), (1.4, 0.2), (1.6, 0.3), (1.8, 0.4), (2, 0.5), (2.2, 0.6) \\ (2.4, 0.7), (2.6, 0.8), (2.8, 0.9), (3, 1.0), (3.2, 0.9), (3.4, 0.8) \\ (3.6, 0.7), (3.8, 0.6), (4, 0.5), (4.2, 0.4), (4.4, 0.3), (4.6, 0.2), (4.8, 0.1) \end{array} \right\}.$$

Note: As we see, the α-cut arithmetic if repeatedly performed in an equation will accumulate the fuzziness of all fuzzy numbers involved. This property can be observed in complex systems when performed for each fuzzy interval. Therefore, in order to reduce fuzzy accumulation, the approximate fuzzy arithmetic operations adopt the weakest t-norm arithmetic operations with different values. The weakest t-norm operations can get more exact performance, which means smaller fuzzy spreads, under uncertain environment.

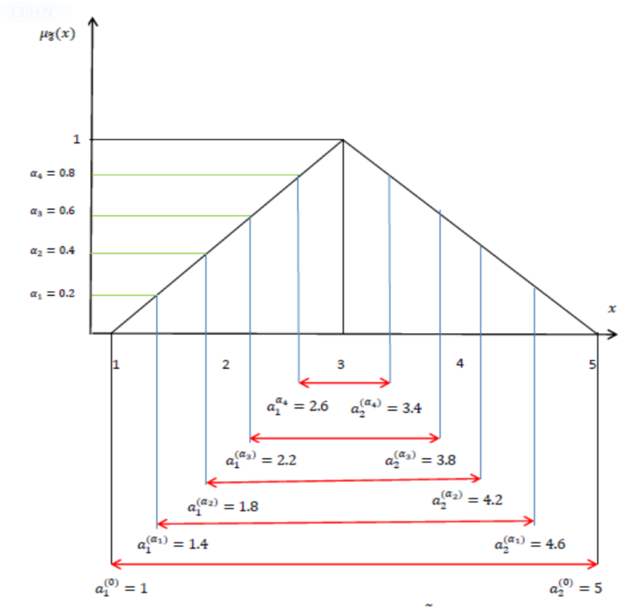

Figure 3: Presentation of the symmetric triangular fuzzy number $\tilde{3}$.

3 The extended LOF algorithm

Regarding the presented definition in Section 2, here, we are going to present an extended LOF algorithm for detecting the fuzzy outliers in a multidimensional database. First of all, we have to introduce the ordinary LOF algorithm for determining the crisp outlier data briefly. To this end, based on [2], the ordinary LOF algorithm is as follows:

Step 1) Introduce minpts parameter and data set D.
Step 2) For any $p \in D$, calculate $minpts - dist(p)$ and $N_{minpts-dist}(p)$ using Definitions 1 and 2.
Step 3) Obtain $reach-dist_{minpts}(p, o)$ (for every $o \in N_{minpts-dist}(p)$ using Definition 3.
Step 4) Calculate $lrd_{minpts}(o)$ (for every $o \in N_{minpts-dist}(p)$ using Definition 4.
Step 5) Calculate LoF_{minpts} coefficient by using Definition 5.
Step 6) Compute LOF coefficient for total data set as p (outlier factor).

Data with a big coefficient is considered as an outlier data (One may compare the obtained factors with a given threshold bound the outliers could be identified).

Now let D be a database including some n-vector with symmetric triangular fuzzy numbers. We know that the LOF coefficient of a symmetric triangular fuzzy data can be obtained by using Definitions 1 to 5; the only difference is in using fuzzy equivalent mathematical concept instead of the used mathematical terms in ordinary LOF; for example, we use triangular fuzzy numbers operations instead of the crisp one. To do this, first, we show any symmetric triangular fuzzy number by different amount of α level with an interval in x axis, and then, we apply fuzzy relations on this interval to allocate an exact coefficient to each interval by left and right scoring method. When LOF coefficients of the data set are computed in this way, by regarding this exact value we follow the LOF procedure in the crisp case to discuss whether it is an outlier or not.

Now, based on the above discussion, we present the following modified LOF algorithm for determining the multidimensional symmetric triangular fuzzy outliers; here $\tilde{p}$ is a data being studied and the one which we want to obtain its inconsistency coefficient, and $\tilde{o}_i$'s are the other data. The main idea of the algorithm is as follows: first by using average weighted method,$\tilde{p}$ is transferred into a one dimensional symmetric fuzzy number; then this number is shown by an α-cuts for given α. Calculating the distance of $\tilde{p}$ to the other $\tilde{o}_i$'s and arranging them by pairwise comparison help us to compute the LOF factor as a fuzzy number. Then, defuzzyfying this number helps us to create the outliers by using the same manner as explained in [2]. The steps of the new algorithm can be presented as follow:

Step 1) Present the associated weight vector and determine the average weight of every fuzzy multidimensional number by using Definition 10.
Step 2) Express fuzzy numbers of step 1 by α-cuts.
Step 3) For any fuzzy number $\tilde{p}$ calculate the distance of the interval $\tilde{p}$ data to the other fuzzy data by using the concept of fuzzy subtraction (Definition 7).
Step 4) Arrange the obtained fuzzy distance from the previous steps by pairwise comparison [6].
Step 5) Calculate the fuzzy LOF factor for every $\tilde{p}$. according to [2] and the given minpts parameter as follow based on Definitions 3 and 4:

$$LOF(\tilde{p}) = \frac{1}{|N_{minpts}(p)|} \times \sum_{o \in N_{minpts}(p)} lrd(o) \times lrd(p)^{-1}.$$

Step 6) Use the left and right scoring defuzzyfication, for every fuzzy LOF factor to allocate an exact amount to each data and then arrange them as an array. Thus, priority is given according to the amount of LOF coefficients, as the mentioned manner for them in [2].

4 Simulations and discussion

To create the general data for testing, initially we produced symmetric fuzzy numbers (l, m, u) randomly; first the mean m is selected randomly by Matlab software in an closed interval. The number β in $(0, 2)$ is selected to produce $(m - \beta, m, m + \beta)$ as a fuzzy randomly produced number. In this manner, by producing 9 of them, a nine-dimensional data is created.

According to the weight conditions $\sum_{i=1}^{n} W_i = 1$ and $W_i \in [0, 1]$, we suppose the weight vector as $W = (\frac{1}{n}, \dots, \frac{1}{n})$, but any weight vector can be selected by decision maker. Thus, a decision maker has the option to give appropriate weight to obtain acceptable results.

Case 1: We examined the above algorithm for a set of 9-dimensional fuzzy data that had been selected randomly by using Matlab 2010 software as explained above; this set was consisted of 100 data in which the mean of 90 of them were selected randomly in the range of 50 to 60. The mean of other 10 data (outliers) were selected in range of 40 to 45 randomly. Results have been recorded for high outlier factor in Table 1 and their defuzzyfied factor are plotted in Figure 4. If we consider Figure 4, we will see that the results are consistent with Table 1. The important fact is that in both cases, the presented algorithm is worked efficiently and defects all the outliers precisely in an easy manner. One can produce these numbers by the following algorithm:

```
Algorithm 1. The Algorithm for the Case 1:
Step 1: Take random numbers $m_{ij} \in (52, 58)$, $i = 1, 2, \dots, 90$, $j = 1, 2, \dots, 9$
Step 2: Take random numbers $m_{ij} \in (42, 43)$, $i = 91, 92, \dots, 100$, $j = 1, 2, \dots, 9$
Step 3: Take random numbers $\beta_{ij} \in (0, 2)$, $i = 1, 2, \dots, 100$, $j = 1, 2, \dots, 9$
Step 4: Take random numbers $w_{ij} \in [0, 1]$ so that for each $i$, $\sum_{j=1}^{9} w_{ij} = 1$ when $i = 1, 2, \dots, 100$, $j = 1, 2, \dots, 9$ (one may put $w_{ij} = \frac{1}{9}$, $i = 1, 2, \dots, 100$, $j = 1, 2, \dots, 9$)
Step 5: Set $\tilde{A}_i = \left(\sum_{j=1}^{9} w_{ij}(m_{ij} - \beta_{ij}), \sum_{j=1}^{9} w_{ij} m_{ij}, \sum_{j=1}^{9} w_{ij}(m_{ij} + \beta_{ij})\right)$, $i = 1, 2, \dots, 100$, $j = 1, 2, \dots, 9$.
```

Case 2: For 100 nine-dimensional data that were selected randomly in the range of 2 to 20, we repeated the test; thus, in this case, the outliers are not predetermined. The results are provided in Table 2 and outlier factor according to the data number is plotted in Figure 5. In fact, for the threshold amount 1.2 the results show that 10 data are the most outlier one. Note that, we select data in range of 40 to 45 randomly. One can produce these numbers by the following algorithm:

Table 1: Weighted average and LOF factor of the 20 last data in case 1.

Number of data	Weighted Average	LOF factor	Prioritize
67	(53.2103,54.1905,55.1707)	(0.0743,1.0075,13.1476)	81
89	(53.1942,54.3964,55.5985)	(0.0742,1.0011,13.2560)	82
37	(53.3093,54.2516,55.1940)	(0.0738,1.0024,13.2604)	83
26	(53.3237,54.3868,55.4498)	(0.0734,1.0067,13.2416)	84
73	(53.2764,54.2595,55.2426)	(0.0739,1.0059,132886)	85
72	(52.8809,54.1850,55.4892)	(0.0744,1.0065,13.2990)	86
75	(53.1554,54.2960,5.4365)	(0.0740,1.0118,13.3508)	87
30	(53.1505,54.2628,55.3751)	(0.0740,1.0134,13.3743)	88
46	(53.0506,54.2381,55.4257)	(0.0737,1.0123,13.4047)	89
52	(52.9373,54.2482,55.5592)	(0.0738,1.0128,13.4906)	90
100	(42.1369,43.1879,44.2389)	(1.2795,2.5281,10.7009)	91
92	(41.9587,43.0742,44.1897)	(1.2833,2.5419,10.7768)	92
99	(41.7388,42.7076,43.6763)	(1.3215,2.5864,10.8724)	93
96	(41.7919,42.5192,43.2466)	(1.3532,2.6093,10.8515)	94
93	(41.6254,42.2969,42.9685)	(1.3739,2.6362,10.9234)	95
95	(4.2202,42.5989,43.9776)	(1.3000,2.5996,11.0918)	96
97	(41.4039,42.3522,43.3004)	(1.3497,2.6295,11.0164)	97
91	(41.0176,42.1822,43.3468)	(1.3468,2.6501,11.1805)	98
94	(40.6102,41.9714,43.3327)	(1.3484,2.6757,11.3539)	99
98	(40.6405,41.8820,43.1234)	(1.3638,2.6866,11.3418)	100

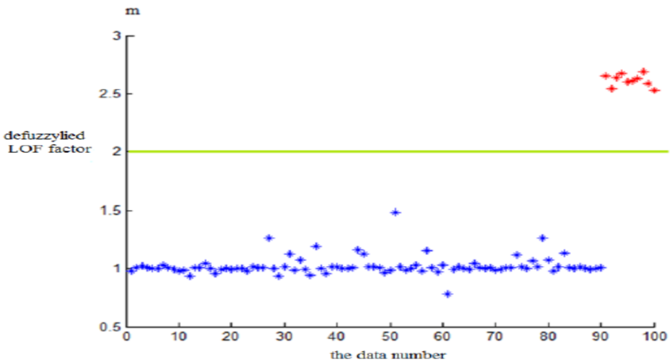

Figure 4: Outlier factors of fuzzy 9 dimensional numbers in case 1.

Table 2: Weighted average and LOF factor of the 20 last data in case 2.

Number of data	Weighted Average	LOF factor	Prioritize
18	(10.2482,11.0734,11.8985)	(0.0742,0.9983,13.2750)	81
47	(9.539,10.9866,12.2194)	(0.0739,1.0086,13.2851)	82
81	(10.1937,11.0458,11.8980)	(0.0743,1.0020,13.3077)	83
32	(10.0244,11.0587,12.0929)	(0.0744,1.0075,13.3077)	84
35	(9.9533,11.1077,12.2621)	(0.0739,1.0075,13.4368)	85
21	(9.8981,11.0948,12.2914)	(0.0741,1.0075,13.4804)	86
29	(9.7612,11.0561,12.3511)	(0.3555,1.2919,109.4159)	87
84	(9.4679,10.8620,12/2562)	(0.0745,1.0063,13.6147)	88
90	(9.7574,11.0791,12.4008)	(0.0743,1.0068,13.6541)	89
56	(9.5082,10.9389,12.3696)	(0.0735,1.0031,13.7013)	90
67	(9.7145,11.0747,12.4349)	(0.0743,1.0061,13.7118)	91
2	(6.0423,6.9917, 7.9411)	(0.0266,0.7186,78.9359)	92
73	(6.0420,6.8094,7.5767)	(0.0670,0.7548,79.8700)	93
88	(7.4460,8.2798,9.1136)	(0.2729,1.1531,103.1452)	94
40	(7.2324,8.2236,9.2147)	(0.2761,1.1749,1052073)	95
65	(4.1071,5.0980,6.0889)	(0.2990,1.2135,107.5364)	96
64	(6.7821,7.6518,8.5216)	(0.3422,1.2781,109.2714)	97
23	(6.7085,7.5261,8.3436)	(0.3555,1.2919,109.4159)	98
53	(6.8233,7.8086,8.7939)	(0.3273,1/2729,110.0384)	99
7	(6.8490,8.1554,9.4617)	(0.2973,1.2424,111.3451)	100

Algorithm 2. The Algorithm for the Case 2:

Step 1: Take random numbers $m_{ij} \in (4, 18),\ i = 1, 2, \ldots, 100,\ j = 1, 2, \ldots, 9$

Step 2: Take random numbers $\beta_{ij} \in (0, 2),\ i = 1, 2, \ldots, 100,\ j = 1, 2, \ldots, 9$

Step 3: Take random numbers $w_{ij} \in [0, 1]$ so that for each i, $\sum_{j=1}^{9} w_{ij} = 1$ when $i = 1, 2, \ldots, 100,\ j = 1, 2, \ldots, 9$ (one may put $w_{ij} = \frac{1}{9},\ i = 1, 2, \ldots, 100,\ j = 1, 2, \ldots, 9$ for simplicity)

Step 4: Set $\tilde{A}_i = \left(\sum_{j=1}^{9} w_{ij}(m_{ij} - \beta_{ij}), \sum_{j=1}^{9} w_{ij} m_{ij}, \sum_{j=1}^{9} w_{ij}(m_{ij} + \beta_{ij})\right)$, $i = 1, 2, \ldots, 100,\ j = 1, 2, \ldots, 9$.

5 Conclusions

LOF is an easy method to apply with achievable results and enough accuracy without passing the complex steps that makes compatible in identifying crises outlier data. Therefore, this method is extended for detecting

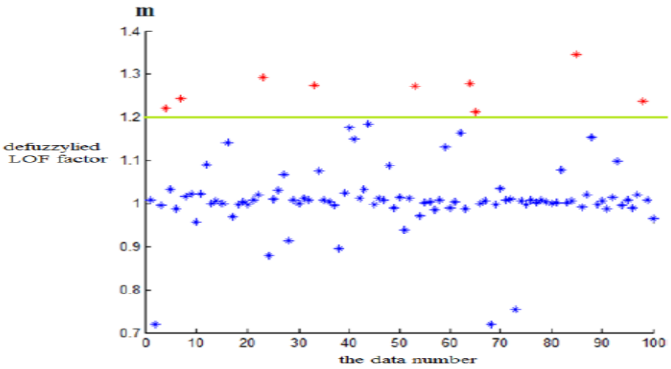

Figure 5: Outlier factors for case 2.

multidimensional fuzzy outliers data. To perform a new algorithm, first LOF factor for each data is calculated; then, by using defuzzyfied left and right scoring, a crisp outlier rate has been calculated for each fuzzy data. Thus by considering a suitable threshold which can be identified by the decision makers, the outliers are detected. More than presenting a new density based method, for detecting fuzzy outliers, a very important advantages of the new method is the fuzzy discipline identification way for high dimensional space. Also based on the testing examples, this algorithm performed efficiently to identify outliers in a directed manner (like case 1) and in undirected and unpredictable manner (like case 2). The obtained theoretical and numerical results show that the presented method is very accurate and successful in determining multidimensional outliers in fuzzy data set.

Some interesting topics for further research could be as: numerical comparison of this algorithm with other existing methods such as LOCI and LOoP, when they extended for fuzzy data, and, studying results of this algorithm on real fuzzy data sets, such as confirmed medical and accounting data.

References

[1] S. H. Abredari, *Fuzzy data processing associated with earthquake precursor*, Research Department of Seismology, Institute of Geophysics, Tehran University, the first conference on earthquake precursor, 2006.

[2] M.M. Breunig, H.P. Krieger, J. Sander, T.R. Ng, *LOF: identifying density based local outliers*, Proc. SIGMOND Conf. **9** (2000) 3–104.

[3] C. Caroni, V. Karioti, *Detecting an innovative outlier in a set of time series*, Comput. Statist. Data Anal. **46** (2004) 561-570.

[4] S. Cateni, V.Colla, G. Nastasi, *A multivariate fuzzy system applied for outliers detection*, J. Intel. Fuzzy Syst. **24** (2013) 889–903.

[5] A.L. Chiu, A.W. Fu, *Enhancements on local outlier detection*, Proc. of the Seventh International Database Engineering and Applications symposium, IEEE, IDEAS, 2003.

[6] D. Dubois, H. Prade, *Fuzzy sets and systems: theory and applications*, Academic Press, New York, 1980.

[7] A. Fakharzadeh J, F. Zarei, *A LoOP based outlier detection method for high dimensional fuzzy data set*, J. Intel. Fuzzy Syst. **32** (2017) 241–248.

[8] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, *Knowledge discovery and data mining: towards a unifying framework*, Proc. 2nd Int. Conf. on Knowledge discovery and Data Mining, Portland, OR, (1996) 82–88.

[9] E.M. Knorr, R.T. Ng, Finding intentional knowledge of distance-based Outliers, Proc. 25th Int. Conf. on Very Large Data Bases, Edinburgh, Scotland: (1998) 211-222.

[10] S. Papadimitriou, H. Kitawaga, P.B. Gibbons, C. Falautsos, *LOCI: Fast outlier detection using the local correlation integral*, Proceedings of the 19th International Conference on Data Engineering (ICDE03), (2003) 315–326.

[11] N. Sadra Abarghouei, H. Hoseini Nasab, A. Sadeghieh and M. Mortazavi, *Fuzzy approach to strategic planning in agriculture section*, Agricultural economics and development, **71** (2010) 179-214 (In Persian).

[12] J.Q. Wang, Z. Zhang, *Aggregation operators on intuitionistic trapezoidal fuzzy number and its application to multi-criteria decision making problems*, Journal of Systems Engineering and Electronics **20**(2) (2009), 321–326.

[13] C.T. Yeh, *A note on trapezoidal approximations of fuzzy numbers*, Fuzzy Sets and Systems, **158** (2007) 747–754.

[14] H. Zimmermann, *Fuzzy set theory and its applications*, Springer, New York, 2001.