Describing Fuzzy Membership Function and Detecting the Outlier by Using Five Number Summary of Data

Md. Farooq Hasan*, Md. Abdus Sobhan

Department of Mathematics and Statistics, Bangladesh University of Business and Technology, Dhaka, Bangladesh
Email: *farooq@bubt.edu.bd, sobhanstatju@gmail.com

Abstract

One of the most important activities in data science is defining a membership function in fuzzy system. Although there are few ways to describe membership function like artificial neural networks, genetic algorithms etc.; they are very complex and time consuming. On the other hand, the presence of outlier in a data set produces deceptive results in the modeling. So it is important to detect and eliminate them to prevent their negative effect on the modeling. This paper describes a new and simple way of constructing fuzzy membership function by using five-number summary of a data set. Five states membership function can be created in this new method. At the same time, if there is any outlier in the data set, it can be detected with the help of this method. Usually box plot is used to identify the outliers of a data set. So along with the new approach, the box plot has also been drawn so that it is understood that the results obtained in the new method are accurate. Several real life examples and their analysis have been discussed with graph to demonstrate the potential of the proposed method. The results obtained show that the proposed method has given good results. In the case of outlier, the proposed method and the box plot method have shown similar results. Primary advantage of this new procedure is that it is not as expensive as neural networks, and genetic algorithms.

Keywords

Fuzzy Set, Membership Function, Five Number Summary, Outliers

1. Introduction

Fuzzy logic is used to describe fuzziness which is characterized by its membership function. Simply membership function represents the degree of truth in

fuzzy logic. Membership function was first introduced by Lotfi A. Zadeh in his paper "Fuzzy sets" [1]. The most common fuzzy membership functions are impulsive fuzzy membership function, triangular fuzzy membership function, trapezoidal fuzzy membership function and Gaussian fuzzy membership function [2]. Membership functions can be defined as a technique to solve practical problems by experience rather than knowledge. Fuzzy logic is used to reduce uncertainty in many sectors such as agriculture [3], medicine [4], power systems [5], production [6], transportation [7] etc. where membership function plays an important role in those implementations. Also it has a great impact on solving various problems in fuzzy mathematics [8] [9].

There are several ways to define membership functions [10] [11]. Some of them are horizontal method of membership estimation, vertical method of membership estimation, pairwise-comparison method of membership function estimation, problem specification based membership determination, membership estimation via fuzzy clustering, artificial neural network, and genetic algorithm [10].

The number of articles is increasing in the field of cognitive systems and artificial intelligence where neural network and reasoning systems like fuzzy system were used. In these cases, the membership functions are generally tuned in a cyclic fashion and are closely tied to their associated rule structure [11]. In the case of Genetic Algorithms, the number of research works is very few [12]. There is huge growth made in the strength of genetic algorithms to find optimum solutions. Both neural network and genetic algorithm approaches to determining membership functions generally make use of associated rules in the knowledge base.

Every method has its own merits and demerits depending on the different scheme. Our main aim is to describe the linguistic variables in a proper way. Membership functions of the fuzzy sets attempt to capture the concepts of linguistic variables. Several membership functions representing linguistic concepts such as low, medium, high, and so on are often employed to define states of a variable. Such a variable is usually called a fuzzy variable [13]. Methods applied to build a membership function are erratic to many reasons because of the terminologies used in defining the membership function varies from man to man. Membership functions can either be chosen by the user arbitrarily, based on the user's experience, or can be designed using machine learning methods (e.g., artificial neural networks, genetic algorithms, etc.). In this paper, a new approach is introduced to formulate the five states fuzzy membership functions by using five number summaries of data. Linear representation of these membership functions is in semi trapezoidal, trapezoidal and triangular shape that will also help to identify the outliers in a data set. All the graphs of membership function are plotted by using MATLAB. The method is very easy to understand and apply in any outlier detecting problem.

Outlier is an observation whose value exceeds the values of other observations in the sample by a large amount, perhaps three or four standard deviations away

from the mean value of all the observations [14]. Many inferential procedures are based on the assumption that the population distribution is normal (a certain type of bell curve) [15]. Even a single extreme outlier in the sample warns the investigator that such procedures may be unreliable, and sometimes the presence of several mild outliers conveys the same message [15]. Hence, researcher should be detected before analyzing data whether their existence of any outlier. The most common way is box plot also called box and whisker plot method [16] [17]. Outliers can also be detected by fuzzy clustering [18], fuzzy discriminant analysis [19]. But these methods are very complex and time consuming.

Therefore, in this paper, a different approach is created to detect outliers by using membership functions of data. Firstly, fuzzy membership functions are constructed by using six points: the lower outer fence, lower inner fence, first quartile, third quartile, upper inner fence and the upper outer fence. In this process five states fuzzy membership functions are developed. Outlier can be detected by the degree of membership of first state and the last state. The results obtained by this process match the results obtained by box plot. The novelty of this method is to simultaneously create membership functions of data and identify if there is any outliers. The procedure is very simple and calculates less than neural networks, and genetic algorithm.

The organization of the remaining parts of the paper is as follows. Section 2 contains the preliminary definition of fuzzy set and membership function. In Section 3, algorithm of five number summary and box plot is discussed. Proposed method will be introduced in Section 4. Implementation of new approach with several real life examples are presented in the Section 5. Conclusion of the paper will be pointed out in the last section.

2. Fuzzy Set and Membership Function

The word "fuzzy" means "vagueness (ambiguity)". Fuzziness occurs when the boundary of a piece of information is not well-defined [13]. For example, words like young, tall, good or high are fuzzy. There is no single quantitative value which defines the term young. For some people, age 25 is young and for others, age 35 is young. The concept young has no clean boundary. Fuzzy set theory is an extension of classical set theory where elements have degree of membership between 0 and 1. In traditional set theory, an element is either in or not in a set A, that is $x \in A$ or $x \notin A$; this kind of set is called a crisp set [2]. A fuzzy set is a set that is characterized by a fuzzy membership function $\mu_A(x) \in [0,1]$. If $\mu_A(x) = 0$, it implies that $x \notin A$. On the other hand, if $\mu_A(x) = 1$ then $x \in A$ [2].

2.1. Definition

Two distinct notations are most commonly employed in the literature to denote membership functions. In one of them, the membership function of a fuzzy set "A" is denoted by $\mu_A(x)$; that is $\mu_A : X \to [0,1]$. In the other one, the membership function is denoted by $A(x)$ and has, of course, the same form

$A : X \rightarrow [0,1]$. A membership function for a fuzzy set "A" on the universe of discourse X is defined as $\mu_A : X \rightarrow [0,1]$, where each element of X is mapped to a value between 0 and 1 [13].

2.2. Definition

The membership function fully defines the fuzzy set. Suppose X is a universal set. Then a fuzzy set "A" can be defined as the set of ordered pairs such that $A = \{(x, \mu_A(x)) : x \in X, \mu_A(x) \in [0,1]\}$ [20]. The nonempty set of objects X is called referential set and [0, 1] (the unit interval) is called valuation set and $\forall x \in X$; $\mu_A(x)$ represents the grade of membership of x [13].

2.3. Example

We consider three fuzzy sets $Y : X \rightarrow [0,1], M : X \rightarrow [0,1]$ and $O : X \rightarrow [0,1]$ that represent the concepts of a young, middle-aged, and old person respectively. A reasonable expression of these concepts by trapezoidal membership functions $Y(x), M(x)$ and $O(x)$ is shown in Figure 1. These functions are defined on the interval $X = [0,80]$, where X is the set of ages of human beings such that

$$\text{Young-aged, } Y(x) = \begin{cases} 1; & x \leq 20 \\ \dfrac{35-x}{15}; & 20 < x < 35 \\ 0; & x \geq 35 \end{cases}$$

$$\text{Middle-aged, } M(x) = \begin{cases} 0; & x \leq 20 \\ \dfrac{x-20}{15}; & 20 < x < 35 \\ 1; & 35 \leq x \leq 45 \\ \dfrac{60-x}{15}; & 45 < x < 60 \\ 0; & x \geq 60 \end{cases}$$

$$\text{Old-aged, } O(x) = \begin{cases} 0; & x \leq 45 \\ \dfrac{x-45}{15}; & 45 < x < 60 \\ 1; & x \geq 60 \end{cases}$$

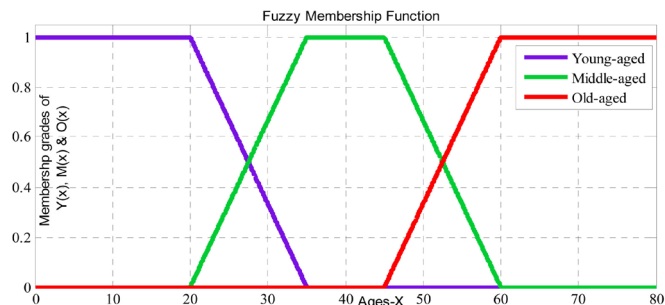Linear representation of these membership functions is given in Figure 1.

Figure 1. Membership functions representing the concepts of a young, middle-aged, and old person.

3. Five Number Summary and Box Plot

The five number summary gives a rough idea about what data set looks like. The five number summary consists of 5 items of a data set [21]. They are:

The minimum,

Q_1 (The first quartile, or 25th percentile),

The median,

Q_3 (The third quartile, or 75th percentile),

The maximum.

The box plot also called box and whisker plot method [16] [17] is a visual representation of a data set based on five number summary [16] [21] or six summaries [22]. It allows for the identification of outliers [15] [23] [24] [25] and [26]. A box plot consists of the following steps [16] [23] [26]:

Step 1: Sort the data on a primary attribute.

Step 2: Calculate the Median, Quartiles Q_1 (25th percentile), Q_3 (75th percentile), and Inter-quartile range: $IQR = Q_3 - Q_1$.

Step 3: Calculate the points that are $1.5 \times IQR$ below Q_1 and $1.5 \times IQR$ above Q_3. These two points are called the lower and the upper inner fences, respectively.

Step 4: A central box extends from the 25th to the 75th percentiles. This box is divided into two compartments at the median value of the data set.

Step 5: The line segments projecting out from the box extend in both directions to the adjacent value. The adjacent values are the points that are 1.5 times the length of the box beyond either quartile. All other data points outside this range are represented individually by little circles; these are considered to be outliers or extreme observations that are not typical of the rest of the data.

The observations that fall outside the two inner fences are called outliers. These outliers can be classified into two kinds of outliers: mild and extreme outliers. To do so, we define two outer fences: a lower outer fence at $3 \times IQR$ below the first quartile and an upper outerfence at $3 \times IQR$ above the third quartile. If an observation is outside either of the two inner fences but within either of the two outer fences, it is called a mild outlier. An observation that is outside either of the two outer fences is called an extreme outlier [26].

4. The New Approach

In a fuzzy membership function, there can be multiple states depending on the domain of the data set. States of the fuzzy variable are fuzzy sets representing linguistic concepts: very low, low, medium, high, very high etc. Now we are going to define five states fuzzy membership functions by using five number summary of a data set discussed in the above section. At first we execute the algorithm of five number summary to produce the preliminary investigation of data set. From summary we will select six points: lower outer fence, lower inner fence, first quartile, third quartile, upper inner fence and upper outer fence. Using these points we construct five states of membership function as follows. We also draw the graph of membership functions by using MATLAB programming

and identify the outliers from this process.

If $X = \{x_1, x_2, x_3, \cdots, x_n\}$ is a set of n observations, and $A_1 : X \rightarrow [0,1]$, $A_2 : X \rightarrow [0,1]$, $A_3 : X \rightarrow [0,1]$, $A_4 : X \rightarrow [0,1]$ and $A_5 : X \rightarrow [0,1]$ are five fuzzy sets defined on X representing the concept of the smallest, small, medium, large and the largest value respectively then the proposed five states fuzzy membership functions are given as follows:

$$\text{Smallest,} \quad A_1(x) = \begin{cases} 1; & x \leq Q_1 - 3I_{qr} \\ \dfrac{Q_1 - (x + 1.5I_{qr})}{1.5I_{qr}}; & Q_1 - 3I_{qr} \leq x \leq Q_1 - 1.5I_{qr} \\ 0; & x \geq Q_1 - 1.5I_{qr} \end{cases}$$

$$\text{Small,} \quad A_2(x) = \begin{cases} 0; & x \leq Q_1 - 3I_{qr} \\ \dfrac{x - (Q_1 - 3I_{qr})}{1.5I_{qr}}; & Q_1 - 3I_{qr} \leq x \leq Q_1 - 1.5I_{qr} \\ \dfrac{Q_1 - x}{1.5I_{qr}}; & Q_1 - 1.5I_{qr} \leq x \leq Q_1 \\ 0; & x \geq Q_1 \end{cases}$$

$$\text{Medium,} \quad A_3(x) = \begin{cases} 0; & x \leq Q_1 - 1.5I_{qr} \\ \dfrac{x - (Q_1 - 1.5I_{qr})}{1.5I_{qr}}; & Q_1 - 1.5I_{qr} \leq x \leq Q_1 \\ 1; & Q_1 \leq x \leq Q_3 \\ \dfrac{(Q_3 + 1.5I_{qr}) - x}{1.5I_{qr}}; & Q_3 \leq x \leq Q_3 + 1.5I_{qr} \\ 0; & x \geq Q_3 + 1.5I_{qr} \end{cases}$$

$$\text{Large,} \quad A_4(x) = \begin{cases} 0; & x \leq Q_3 \\ \dfrac{x - Q_3}{1.5I_{qr}}; & Q_3 \leq x \leq Q_3 + 1.5I_{qr} \\ \dfrac{(Q_3 + 3I_{qr}) - x}{1.5I_{qr}}; & Q_3 + 1.5I_{qr} \leq x \leq Q_3 + 3I_{qr} \\ 0; & x \geq Q_3 + 3I_{qr} \end{cases}$$

$$\text{Largest,} \quad A_5(x) = \begin{cases} 0; & x \leq Q_3 + 1.5I_{qr} \\ \dfrac{x - (Q_3 + 1.5I_{qr})}{1.5I_{qr}}; & Q_3 + 1.5I_{qr} \leq x \leq Q_3 + 3I_{qr} \\ 1; & x \geq Q_3 + 3I_{qr} \end{cases}$$

Where Q_1 = First quartile, Q_3 = Third quartile, and I_{qr} = Inter-quartile range.

The defined fuzzy membership function follows two properties [13] [20]:

- Cross over point: for all states cross over point is 0.5 which indicates these sets are symmetric.
- Height of the fuzzy set is $\max\{\mu_A(x)\} = 1$.

Membership functions allow us to graphically represent a fuzzy set. The x axis

represents the universe of discourse, whereas the y axis represents the degrees of membership in $[0,1]$ interval. Graphs of these functions have trapezoidal, semi trapezoidal and triangular shapes which are most common in current applications. For the outliers we consider the membership functions $A_1(x)$ and $A_5(x)$ only. As outliers are the extremely smallest or the extremely largest value of a data set so we define following two conditions for outliers:

- $\forall x \in X$; If $A_1(x) = 1$ or $A_5(x) = 1$ then x is an outlier.
- $\forall x \in X$; If $A_1(x) < 1$ or $A_5(x) < 1$ then x is a mild outlier.

5. Implementation

There are three real life problems that have been discussed in this section to show the capability of the proposed method. To evaluate the proposed method, real world data sets are used. The list of variables used in this section is given in Table 1. Both even and odd numbers of data sets are taken from Bangladesh University of Business and Technology, and Bangladesh Meteorological Department [27]. At first the five number summary of the data set will be extracted then fuzzy membership functions of the data will be defined using these. Outliers will be identified by plotting membership function on the graph. Box plots are also drawn for each problem to show that the results of the proposed method are effective for outliers.

5.1. Height of People

Below is a list of the height (in centimeter) of 42 male students of the first year 2020 of the Bangladesh University of Business and technology:

170.18, 179.83, 161.38, 172.72, 175.26, 175.26, 170.18, 170.18, 182.88, 173.00, 152.40, 170.69, 179.83, 167.64, 175.26, 172.72, 163.83, 172.72, 170.18, 201.17, 177.80, 160.02, 152.40, 172.72, 177.80, 167.64, 167.64, 160.02, 172.72, 155.45, 170.18, 181.61, 167.64, 142.54, 165.10, 175.26, 167.64, 172.72, 175.26, 177.80, 167.64, 172.72.

Our aim is to investigate the shortest, medium and the tallest students from this data set by using the new approach. Here the number of observations, $n = 42$ is even. The five number summary of these given data is estimated by usual statistical procedure [16] [23] [26]. We get,

The minimum value, $x_{\min} = 142.54$

First quartile, $Q_1 = 167.64$

Median, $Q_2 = 171.71$

Table 1. List of variables used for the evaluation of proposed method.

SN.	Variables	Symbol used for membership functions
1.	Height of people	$H_i(x); i = 1, 2, 3, 4, 5$
2.	Weight of people	$W_i(x); i = 1, 2, 3, 4, 5$
3.	The daily temperature in a city	$T_i(x); i = 1, 2, 3, 4, 5$

Third quartile, $Q_3 = 175.26$

The maximum value, $x_{\max} = 201.17$

Inter Quartile Range, $I_{qr} = Q_3 - Q_1 = 175.26 - 167.64 = 7.62$

Inner fence: $Q_1 - 1.5I_{qr} = 156.21$ and $Q_3 + 1.5I_{qr} = 186.69$.

Outer fence: $Q_1 - 3I_{qr} = 144.78$ and $Q_3 + 3I_{qr} = 198.12$.

The visual representation of this summary pointing with outliers is given in Figure 2.

Now $\forall x \in X$; where X is the data set of heights of the given students, if $H_1 : X \to [0,1]$, $H_2 : X \to [0,1]$, $H_3 : X \to [0,1]$, $H_4 : X \to [0,1]$ and $H_5 : X \to [0,1]$ represent the fuzzy sets of the shortest, short, medium, tall and the tallest students respectively; then by using our proposed method the five states fuzzy membership functions are given as follows:

$$\text{Shortest, } H_1(x) = \begin{cases} 1; & x \leq 144.78 \\ \dfrac{156.21 - x}{11.43}; & 144.78 \leq x \leq 156.21 \\ 0; & x \geq 156.21 \end{cases}$$

$$\text{Short, } H_2(x) = \begin{cases} 0; & x \leq 144.78 \\ \dfrac{x - 144.78}{11.43}; & 144.78 \leq x \leq 156.21 \\ \dfrac{167.64 - x}{11.43}; & 156.21 \leq x \leq 167.64 \\ 0; & x \geq 167.64 \end{cases}$$

$$\text{Medium, } H_3(x) = \begin{cases} 0; & x \leq 156.21 \\ \dfrac{x - 156.21}{11.43}; & 156.21 \leq x \leq 167.64 \\ 1; & 167.64 \leq x \leq 175.26 \\ \dfrac{186.79 - x}{11.43}; & 175.26 \leq x \leq 186.79 \\ 0; & x \geq 186.79 \end{cases}$$

$$\text{Tall, } H_4(x) = \begin{cases} 0; & x \leq 175.26 \\ \dfrac{x - 175.26}{11.43}; & 175.26 \leq x \leq 186.79 \\ \dfrac{198.12 - x}{11.43}; & 186.69 \leq x \leq 198.12 \\ 0; & x \geq 198.12 \end{cases}$$

$$\text{Tallest, } H_5(x) = \begin{cases} 0; & x \leq 186.69 \\ \dfrac{x - 186.69}{11.43}; & 186.69 \leq x \leq 198.12 \\ 1; & x \geq 198.12 \end{cases}$$
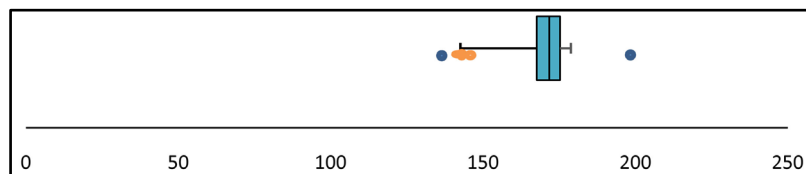

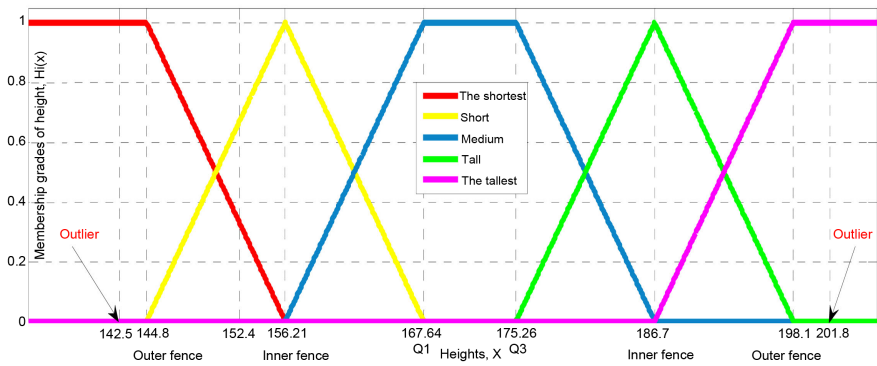Figure 2. Box plot for height of the students.

Figure 3. Membership functions for height of the students and the outliers.

By using MATLAB the graphical representation of these membership functions indicating with the outliers is shown in **Figure 3**.

The box plot shows that there are two outliers in the data set X and they are 142.54 and 201.17. The membership grades of these two data are: $H_1(142.54) = 1$ and $H_5(201.17) = 1$. So both of these are outliers in terms of our condition. On the other hand, box plot shows that there is a suspected outlier at $x = 152.4$ and membership grade of it is $H_1(152.4) = 0.33 < 1$, which meets the conditions for being a mild outlier. Thus the new method of detecting the outlier shows the same result as the box plot.

5.2. Weight of People

Below is a list of the weight (in lbs.) of 15 female students of fifth semester in the year 2020 of the Bangladesh University of Business and technology:

105.8, 187.4, 132.3, 132.3, 105.8, 127.4, 114.6, 99.2, 88.2, 119.0, 110.2, 125.7, 108.0, 143.3, 108.0.

We need to select the thinnest students for a 100 meter running competition. For this, we apply our new approach to construct a five states membership functions of the thinnest, thin, average, heavy and the heaviest students. At first we arrange the data in ascending order of magnitude then the set of observations are:

$$X = \{88.2, 99.2, 105.8, 105.8, 108.0, 108.0, 110.2, 114.6,$$
$$119.0, 125.7, 127.4, 132.3, 132.3, 143.3, 187.4\}$$

The five number summary of these given data is estimated by usual statistical procedure [16] [23] [26]. We get,

The minimum value, $x_{min} = 88.2$

First quartile, $Q_1 = 105.8$

Median, $Q_2 = 114.6$

Third quartile, $Q_3 = 132.3$

The maximum value, $x_{max} = 187.4$

Inter Quartile Range, $I_{qr} = Q_3 - Q_1 = 132.3 - 105.8 = 26.5$

Inner fence: $Q_1 - 1.5I_{qr} = 66.05$ and $Q_3 + 1.5I_{qr} = 172.05$.

Outer fence: $Q_1 - 3I_{qr} = 26.3$ and $Q_3 + 3I_{qr} = 211.8$.

The visual representation of this summary pointing with outliers is given in **Figure 4**.

Now $\forall x \in X$; where X represents the weights of the given female students, if $W_1 : X \to [0,1]$, $W_2 : X \to [0,1]$, $W_3 : X \to [0,1]$, $W_4 : X \to [0,1]$ and $W_5 : X \to [0,1]$ are the fuzzy sets of the thinnest, thin, average, heavy and the heaviest students respectively; then by using our proposed method the five states fuzzy membership functions are given as follows:

$$\text{Thinnest,} \quad W_1(x) = \begin{cases} 1; & x \le 26.3 \\ \dfrac{66.05 - x}{39.75}; & 26.3 \le x \le 66.05 \\ 0; & x \ge 66.05 \end{cases}$$

$$\text{Thin,} \quad W_2(x) = \begin{cases} 0; & x \le 26.3 \\ \dfrac{x - 26.3}{39.75}; & 26.3 \le x \le 66.05 \\ \dfrac{105.8 - x}{39.75}; & 66.05 \le x \le 105.8 \\ 0; & x \ge 105.8 \end{cases}$$

$$\text{Average,} \quad W_3(x) = \begin{cases} 0; & x \le 66.05 \\ \dfrac{x - 66.05}{39.75}; & 66.05 \le x \le 105.8 \\ 1; & 105.8 \le x \le 132.3 \\ \dfrac{172.05 - x}{39.75}; & 132.3 \le x \le 172.05 \\ 0; & x \ge 172.05 \end{cases}$$

$$\text{Heavy,} \quad W_4(x) = \begin{cases} 0; & x \le 132.3 \\ \dfrac{x - 132.3}{39.75}; & 132.3 \le x \le 172.05 \\ \dfrac{211.8 - x}{39.75}; & 172.05 \le x \le 211.8 \\ 0; & x \ge 211.8 \end{cases}$$

$$\text{Heaviest,} \quad W_5(x) = \begin{cases} 0; & x \le 172.05 \\ \dfrac{x - 172.05}{39.75}; & 172.05 \le x \le 211.8 \\ 1; & x \ge 211.8 \end{cases}$$

By using MATLAB the graphical representation of these membership functions indicating with the outliers is shown in **Figure 5**.
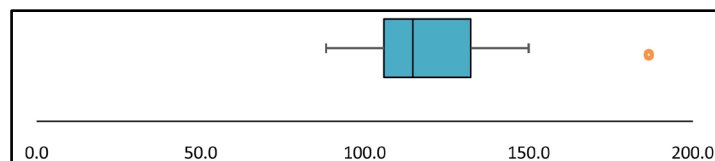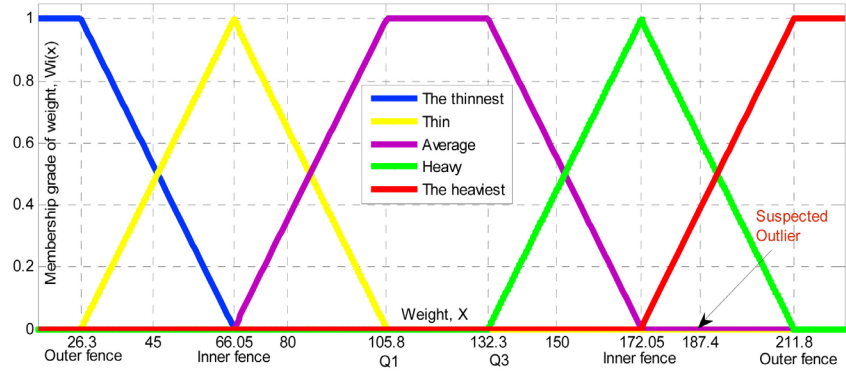

Figure 4. Box plot for weight of the students.

Figure 5. Membership functions for weight of the students.

The box plot shows that there is a suspected outlier at $x = 187.40$ and the membership grade of it is $W_5(187.40) = 0.39 < 1$, which meets the conditions for being a mild outlier. So here too the new method and the box plot are showing the same result.

5.3. Intensity of Daily Temperature

The intensity of heat is not same for all objects. Water begins to boil at a temperature of 100˚C (Celsius) where the melting point of uranium is 1132˚C. So sometimes it is difficult to determine the warmest day or the coldest day of a month according to the temperature. But membership function can reduce the complexity of this problem.

The daily maximum temperature (in ˚C) for the month of May 2020 of Dhaka city is given below [27]:

28, 30, 31, 34, 35, 30, 33, 33, 35, 36, 35, 36, 33, 34, 34, 35, 35, 37, 33, 29, 29, 32, 34, 33, 33, 32, 29, 31, 33, 32, 29.

Using the new approach, we want to check whether the days of May are warm or temperate according to the temperature. At first we arrange the data in ascending order of magnitude then the set of observations are:

$$X = \{28, 29, 29, 29, 29, 30, 30, 31, 31, 32, 32, 32, 33, 33, 33, 33, 33, 33, 33,$$

$$34, 34, 34, 34, 35, 35, 35, 35, 35, 36, 36, 37\}$$

Here the number of observations, $n = 31$ is odd. The five number summary of these data is estimated by usual statistical procedure [16] [23] [26]. We get,

The minimum value, $x_{\min} = 28$

First quartile, $Q_1 = 31$

Median, $Q_2 = 33$

Third quartile, $Q_3 = 35$

The maximum value, $x_{\max} = 37$

Inter Quartile Range, $I_{qr} = Q_3 - Q_1 = 35 - 31 = 4$

Inner fence: $Q_1 - 1.5I_{qr} = 25$ and $Q_3 + 1.5I_{qr} = 41$.

Outer fence: $Q_1 - 3I_{qr} = 19$ and $Q_3 + 3I_{qr} = 47$.

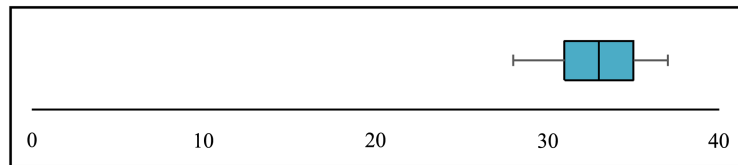The visual representation of this summary is given in **Figure 6**.

Figure 6. Box plot of daily temperature.

Now if we define fuzzy sets $T_1 : X \to [0,1]$, $T_2 : X \to [0,1]$, $T_3 : X \to [0,1]$, $T_4 : X \to [0,1]$ and $T_5 : X \to [0,1]$ representing the temperature of very low, low, medium, high and very high respectively, then $\forall x \in X$; the five states membership functions are given as follows:

$$\text{Very low, } T_1(x) = \begin{cases} 1; & x \leq 19 \\ \dfrac{25-x}{6}; & 19 \leq x \leq 25 \\ 0; & x \geq 25 \end{cases}$$

$$\text{Low, } T_2(x) = \begin{cases} 0; & x \leq 19 \\ \dfrac{x-19}{6}; & 19 \leq x \leq 25 \\ \dfrac{31-x}{6}; & 25 \leq x \leq 31 \\ 0; & x \geq 31 \end{cases}$$

$$\text{Medium, } T_3(x) = \begin{cases} 0; & x \leq 25 \\ \dfrac{x-25}{6}; & 25 \leq x \leq 31 \\ 1; & 31 \leq x \leq 35 \\ \dfrac{41-x}{6}; & 35 \leq x \leq 41 \\ 0; & x \geq 41 \end{cases}$$

$$\text{High, } T_4(x) = \begin{cases} 0; & x \leq 35 \\ \dfrac{x-35}{6}; & 35 \leq x \leq 41 \\ \dfrac{47-x}{6}; & 41 \leq x \leq 47 \\ 0; & x \geq 47 \end{cases}$$

$$\text{Very high, } T_5(x) = \begin{cases} 0; & x \leq 41 \\ \dfrac{x-41}{6}; & 41 \leq x \leq 47 \\ 1; & x \geq 47 \end{cases}$$

By using MATLAB the graphical representation of these membership functions is shown in **Figure 7**.

Graph of membership function does not display any outliers that box plot also didn't. This means that in May 2020, there was not a day in Dhaka city where the temperature was excessive high or too low.
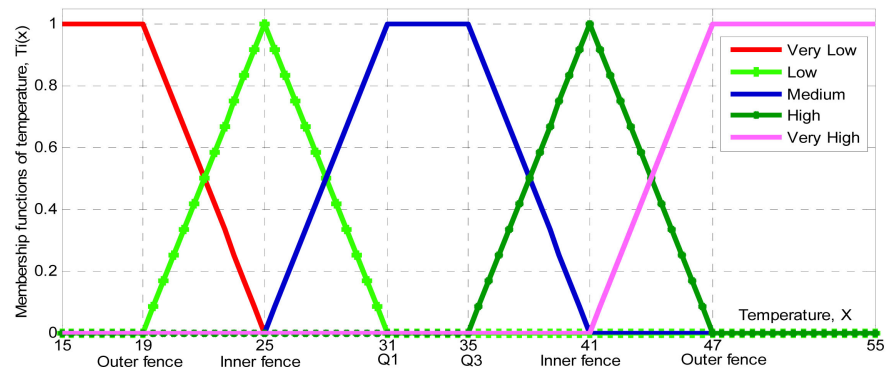
Figure 7. Membership function for daily temperature.

6. Conclusion

An efficient method for generating membership functions and identifying the outliers is discussed in this paper. By applying it, the solution of some real life related problems has also been highlighted. Each problem is presented graphically and their results are also described. The results obtained are excellent and provide adequate information of the data set. The set of height and weight of human has been successfully defined by using it. Linguistic variables can be easily captured by defining their membership function in this process. Five states membership functions of linguistic variables such as tiny, short, medium, tall and giant can be identified by this approach. In this procedure triangular, semi-trapezoidal and trapezoidal shaped membership functions have been constructed; those play an important role in the identification of outliers. There is no difference between the outliers obtained by this process and the outliers obtained in the box plot method. So this new method is effective for simultaneously creating membership functions of data and detecting outliers. This proposed method is also less computationally, less expensive and time saving than the most common methods such as neural networks, fuzzy clustering, and genetic algorithms. Though triangular and trapezoidal membership functions have been described in this paper but there are other membership functions such as sigmoidal, Gaussian, z-shape, s-shape functions. Our future aim is to define these types of functions using the proposed method.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zadeh, L.A. (1965) Fuzzy Sets. *Information and Control*, **8**, 338-353.
     https://doi.org/10.1016/S0019-9958(65)90241-X

[2] Ling, W.K. (2007) Nonlinear Digital Filters: Analysis and Applications. Academic Press, Cambridge.

[3] Kurniasih, D., Jasmi, K.A., Basiron, B., Huda, M. and Maseleno, A. (2018) The Uses

of Fuzzy Logic Method for Finding Agriculture and Livestock Value of Potential Village. *International Journal of Engineering & Technology*, **7**, 1091-1095. https://doi.org/10.14419/ijet.v7i3.11984

[4] Phuong, N.H. and Kreinovich, V. (2001) Fuzzy Logic and Its Applications in Medicine. *International Journal of Medical Informatics*, **62**, 165-173. https://doi.org/10.1016/S1386-5056(01)00160-5

[5] Handschin, E., Hoffmann, W., Reyer, F., Stephanblome, T., Schlucking, U., Westermann, D. and Ahmed, S.S. (1994) A New Method of Excitation Control Based on Fuzzy Set Theory. *IEEE Transactions on Power Systems*, **9**, 533-539. https://doi.org/10.1109/59.317569

[6] Zhang, J., Su, Y., Wu, J. and Liang, H. (2015) GIS Based Land Suitability Assessment for Tobacco Production Using AHP and Fuzzy Set in Shandong Province of China. *Computers and Electronics in Agriculture*, **114**, 202-211. https://doi.org/10.1016/j.compag.2015.04.004

[7] Teodorović, D. (1994) Fuzzy Sets Theory Applications in Traffic and Transportation. *European Journal of Operational Research*, **74**, 379-390. https://doi.org/10.1016/0377-2217(94)90218-6

[8] Hasan, M.F., Badsha, M.F. and Sobhan, M.A. (2019) A Different Approach to Solve Fuzzy Quadratic Equation $AX^2 = B$. *IOSR Journal of Engineering* (*IOSRJEN*), **9**, 1-7.

[9] Hasan, M.F. and Sultana, A. (2017) A New Approach to Solve Fuzzy Linear Equation $AX + B = C$. *IOSR Journal of Mathematics* (*IOSR-JM*), **13**, 22-30.

[10] George, J.K. and Bo, Y. (1995) Fuzzy Sets and Fuzzy Logic: Theory and Applications. PHI New, Delhi, 443-455.

[11] Hayashi, I., Nomura, H., Yamasaki, H. and Wakami, N. (1992) Construction of Fuzzy Inference Rules by NDF and NDFL. *International Journal of Approximate Reasoning*, **6**, 241-266. https://doi.org/10.1016/0888-613X(92)90019-V

[12] Karr, C.L. and Gentry, E.J. (1993) Fuzzy Control of pH Using Genetic Algorithms. *IEEE Transactions on Fuzzy Systems*, **1**, 46. https://doi.org/10.1109/TFUZZ.1993.390283

[13] Klir, G. and Yuan, B. (1995) Fuzzy Sets and Fuzzy Logic (Vol. 4). Prentice Hall, Upper Saddle River. https://doi.org/10.1109/45.468220

[14] Gujarati, D.N. (2009) Basic Econometrics. Tata McGraw-Hill Education, New York.

[15] Devore, J.L. (2015) Probability and Statistics for Engineering and the Sciences. Cengage Learning, Boston.

[16] Thirumalai, C., Vignesh, M. and Balaji, R. (2017) Data Analysis Using Box and Whisker Plot for Lung Cancer. 2017 *Innovations in Power and Advanced Computing Technologies*, Vellore, 21-22 April 2017, 1-6. https://doi.org/10.1109/IPACT.2017.8245071

[17] Vignesh, V., Pavithra, D., Dinakaran, K. and Thirumalai, C. (2017) Data Analysis Using Box and Whisker Plot for Stationary Shop Analysis. 2017 *International Conference on Trends in Electronics and Informatics*, Tirunelveli, 11-12 May 2017, 1072-1076. https://doi.org/10.1109/ICOEI.2017.8300874

[18] Moh'd Belal, A.Z., Al-Dahoud, A. and Yahya, A.A. (2010) New Outlier Detection Method Based on Fuzzy Clustering. *WSEAS Transactions on Information Science and Applications*, **7**, 681-690.

[19] Lin, C.C. and Chen, A.P. (2004) Fuzzy Discriminant Analysis with Outlier Detection by Genetic Algorithm. *Computers & Operations Research*, **31**, 877-888.

https://doi.org/10.1016/S0305-0548(03)00040-6

[20] Zimmermann, H.J. (2011) Fuzzy Set Theory—And Its Applications. Springer Science & Business Media, Berlin.

[21] Baron, M. (2019) Probability and Statistics for Computer Scientists. CRC Press, Boca Raton. https://doi.org/10.1201/9781315172286

[22] Pranatha, M.D.A., Pramaita, N., Sudarma, M. and Widyantara, I.M.O. (2018) Filtering Outlier Data Using Box Whisker Plot Method for Fuzzy Time Series Rainfall Forecasting. 2018 4*th International Conference on Wireless and Telematics*, Bali, 12-13 July 2018, 1-4. https://doi.org/10.1109/ICWT.2018.8527734

[23] Le, C.T. and Eberly, L.E. (2016) Introductory Biostatistics. John Wiley & Sons, Hoboken.

[24] Larsen, R.D. (1985) Box-and-Whisker Plots. *Journal of Chemical Education*, **62**, 302. https://doi.org/10.1021/ed062p302

[25] Schwertman, N.C., Owens, M.A. and Adnan, R. (2004) A Simple More General Box Plot Method for Identifying Outliers. *Computational Statistics & Data Analysis*, **47**, 165-174. https://doi.org/10.1016/j.csda.2003.10.012

[26] Mann, P.S. (2007) Introductory Statistics. John Wiley & Sons, Hoboken.

[27] Bangladesh Meteorological Department. http://bmd.gov.bd