Interval forecasting model for time series based on the fuzzy clustering technique

T Vovan¹ and D Phamtoan²,3,4*

¹College of Nature Sciences, Can Tho University
²University of Science, Ho Chi Minh City, Vietnam
³Vietnam National University, Ho Chi Minh City, Vietnam
⁴Faculty of Engineering, Van Lang University, Ho Chi Minh City, Vietnam

*dinh.pt@vlu.edu.vn

Abstract. This paper proposes the forecasting model for the fuzzy time series based on the improvement of the background data and fuzzy relationship (IFTC). This algorithm is built based on the fuzzy cluster analysis which the suitable number of clusters for series is considered. The problem of interpolating data according to fuzzy relationships of time series in the trapezoidal fuzzy number is also established. The proposed model is illustrated step by step by a numerical example and effectively implemented by the Matlab procedure. The IFTC has advantages in comparing to other models via the several indexes such as the MAE, MAPE and MSE with the Enrollment dataset.

1. Introduction

Forecasting is the process of making predictions based on historical data, knowledge and experience of the related problems. Because of its important role in many fields, forecasting has been paying much attention by scientists. It is an important science basis for projects, policies and appropriate development strategy. Thus, forecasting is always interested in managers and scientists. Regarding data, time series is popular and has great demand on forecasting in reality. For this data, the two main models used for forecasting are regression and time series. The regression model has conditional constraints on data that are difficult to satisfy in reality, so it has the disadvantage in many cases [5,8]. The time series model (TS) is considered to have many more advantages, so it is used very popularly today. Many scientists utilized the TS models as Auto regression (AR), Autoregressive Intergrated Moving Average (ARIMA) to apply in Economy, environment and hydrology [7,13,17].

For the mechanical and manufacturing engineering field, there are some outstanding studies as Fang-mei tseng et al. [21] proposed a procedure of fuzzy seasonal time series and apply this method to forecasting the production value of the mechanical industry in Taiwan. This method includes interval models with interval parameters and provides the possibility distribution of future value. From the results of practical application to the mechanical industry, it can be shown that this method makes good forecasts. Further, this method makes it possible for decision makers to forecast the possible situations.
based on fewer observations than the SARIMA model and has the basis of pre-procedure for fuzzy time series. Trendafilova [22] considered some possibilities to use pure time series analysis for damage diagnosis in vibrating structures. The author introduced the basics of the state space methodology and discussed a number of possible methods to extract damage sensitive features from the state space representation of the attractor of a vibrating system. In similar direction, Irina Trendafilova and Emil Manoach [20] introduced two viable VHM (Vibration health monitoring) methods that use large amplitude vibrations and are based on nonlinear time series analysis and suggested explore some changes in the state space geometry/distribution of the structural dynamic response with damage and their use for damage detection purposes. However, to build a best TS model, the time series must stop and its error must be the white noise. These conditions are considered the challenge that are difficult to satisfy in reality. Therefore, in many cases, the forecasting results are poor when using TS models. Although, many authors as Abreu et al. [2], Ghosh, Chowdhury and Prajneshu [12], Tai [19], Yu and Huarng [23] have improved the previous model, they still face many difficulties.

One of the research directions that has been interested by many statisticians is defuzzification for the original data to get the relationship of the elements in series. Song and Chissom [18] are pioneers in this direction. The TS models normally include three phases: (i) determine the background set from the original data, divide the interval for the background set, and find the number of elements for each interval; (ii) build the fuzzy relationship, and (iii) interpolate and predict the time series. For (i), many authors have used the minimum and maximum values of the original data to define the background set [7,8]. Moreover, Huarng [13], Huarng and Yu [23] proposed the two new techniques for determining intervals of the background set based on the mean and distribution of the entire series. Another way to build the background set use the clustering method. This method is a new research direction at present. For (ii), some researchers have performed this problem, for instance, Song and Chissom [18] used the matrix operations and Chen [8] utilized the fuzzy relationship. Meanwhile, Aladag [3] used the Neural Network to define the fuzzy relationship. For (iii), most of study applied the centroid method in order to perform this phase [8,13,20].

This paper contributes to the three stages (i),(ii) and (iii) for the TS model. For (i), after normalizing the data, the model find the number of groups that belong to each centroid of the time series. These centroids are calculated according to the algorithm to determine the suitable number of clusters (SNC). For (ii), based on the FCM, we build the algorithm to find the fuzzy relationship between the centroid and each value of the time series by interval after that perform the process of interpolating according to research of Liu Hao-Tien [16]. This is the contribution to stage (iii). Combining all the enhancements, this article proposes an interpolating model for time series that is better than the existing models based on the benchmark dataset.

The remainder of this paper is structured as follows: Section 2 considers some conceptions about time series, fuzzy number related to the proposed model. Section 3 presents the step by step of the developed algorithm. Section 4 illustrates the proposed algorithm and compares to other ones based on the benchmark dataset. The final section is the conclusion.
2. Some conception

**Definition 1.** Give \( U \) is universe set, \( U = \{u_1, u_2, \ldots, u_n\} \). Fuzzy set \( A \) of \( U \) is defined as follows:

\[
A = \{ \mu_A(u_1) / u_1, \mu_A(u_2) / u_2, \ldots, \mu_A(u_n) / u_n \},
\]

where \( \mu_A \) is the membership function \( A \), \( \mu_A : U \rightarrow [0, 1] \). \( \mu_A(u_i) \) indicates the grade of membership of \( u_i \) in \( A \), \( \mu_A(u_i) \in [0, 1], 1 \leq i \leq n \).

**Definition 2.** Suppose \( F(t) \) is computed from \( F(t-1) \), the fuzzy logical relationship between \( F(t) \) and \( F(t-1) \) is illustrated by the following formula:

\[
F(t) = F(t-1) \ast R(t, t-1),
\]

where, \( \ast \) is an arithmetic operator, \( R(t, t-1) \) is the fuzzy logical relationship. If \( F(t-1) = A_{t-1} \) and \( F(t) = A_t \) then the fuzzy logical relationship is signed as \( A_{t-1} \rightarrow A_t \).

**Definition 3.** Given a series of original data \( \{X_i\} \) and the predictive value \( \{\hat{X}_i\}, i = 1, 2, \ldots, n \). Then, we have some parameters as follows:

- Mean squared error: \( MSE = \frac{1}{n} \sum_{i=1}^{n} (\hat{X}_i - X_i)^2 \).
- Mean absolute error: \( MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{X}_i - X_i}{X_i} \right| \).
- Mean absolute percentage error: \( MAPE = \frac{1}{n} \sum_{i=1}^{n} \left\{ \left| \frac{\hat{X}_i - X_i}{X_i} \right| \times 100 \right\} \).

The value of these parameters are much smaller, then the proposed model is the best.

**Definition 4:** A trapezoidal fuzzy number \( A^* \) can be defined as follows: \( A^*=(a,b,c,d) \).

\[
\mu_{A^*}(x) = \begin{cases} 
0, & x < a \text{ or } x > d \\
\frac{x-a}{b-a}, & a \leq x \leq b \\
1, & b \leq x \leq c \\
\frac{d-x}{d-c}, & c \leq x \leq d 
\end{cases}
\]

3. The proposed algorithm

Let \( X_i, i = 1, N \) is time series. The proposed fuzzy time series model (FTSF) includes the seven steps as follows:

**Step 1.** Normalize the time series to scale 100:

\[
Y_i = 100X_i / \max \{X_i\}, i = 1, 2, \ldots, N.
\]
Step 2. Determine the suitable number of clusters for the universal set $U$. This problem is performed by the SNC (Suitable number of clusters) algorithm

Step 2.1. When $t = 0$ (the number of current iterations is zero), initialize

$$V^{(0)} = \{V_1^{(0)}, V_2^{(0)}, \ldots, V_N^{(0)}\} = X = \{Y_1, Y_2, \ldots, Y_N\}, \; \varepsilon > 0$$

is a very small number.

Step 2.2. Update the sequence of centroids by the formula below:

$$V_i^{(t+1)} = \frac{\sum_{j=1}^{N} f\left(V_i^{(t)}, V_j^{(t)}\right) V_j^{(t)}}{\sum_{j=1}^{N} f\left(V_i^{(t)}, V_j^{(t)}\right)}, \; i = 1, \ldots, N,$$

where

$$f\left(V_i^{(t)}, V_j^{(t)}\right) = \begin{cases} e^{-\frac{d(V_i^{(t)}, V_j^{(t)})^2}{\lambda}} & \text{if } d(V_i^{(t)}, V_j^{(t)}) \leq c_t, \\ 0 & \text{if } d(V_i^{(t)}, V_j^{(t)}) > c_t \end{cases}$$

with $d(V_i^{(t)}, V_j^{(t)})$ is Euclidean distance between $V_i^{(t)}$ and $V_j^{(t)}$. $c_t = \frac{1}{n(n-1)} \sum_{i<j} d(V_i, V_j)$ is the average of the similar index of cluster of all pairs of data and $\lambda = \frac{d}{r}$. In fact, the window size $\lambda$ determines the number of clusters in the time series, if $\lambda \to 0$, each element will belong to their own cluster, by contrast, all elements of the time series will belong to one cluster ($\lambda \to \infty$). The value of $\lambda$ depends on constant $r$ and $c_t$. In the numerical example of this articles, we take $\lambda = d/5$.

Step 2.3. Repeat Step 2 until $\max_i c\left(V_i^{(t+1)}, V_i^{(t)}\right) < \varepsilon$.

In this algorithm, after each iteration, each $V_i^{(t)}$ converges to the centroid of cluster containing it. This process stops when the variation of all $V_i^{(t)}$ through two of adjacent iterations are less than $\varepsilon$. When $\varepsilon$ is large, the algorithm will stop faster but the number of cluster can be unsuitable In this article, we choose $\varepsilon = 0.0001$ for illustrative example.

Step 3. Determine the specific elements in each cluster by the Fuzzy Cluster Analysis (FCA) algorithm with four steps as follows:

Step 3.1. Divide $Y_i, i=1,N$ into $k$ intervals $C_1, C_2, \ldots, C_k$ randomly. Establish the initial partition matrix $U^{(0)} = [\mu_{ij}]_{k \times n}$, with $\mu_{ij} = 1$ if the $j^{th}$ element belongs to the $i^{th}$ cluster and $\mu_{ij} = 0$ for otherwise.

Step 3.2. Find the representative element $v_i$ for each population by the formula below:

$$v_i = \left( \sum_{j=1}^{n} \mu_{ij}^2 Y_j \right) / \left( \sum_{j=1}^{n} \mu_{ij}^2 \right),$$

where $1 \leq i \leq k$, $\mu_{ij}$ is the probability of $j^{th}$ element assigned to $C_i$.
Step 3.3. Update the new partition matrix $U^{(i)}$ by the following principle:

$$
\mu_{ij}^{(i)} = \frac{1}{\sum_{l=1}^{k} [d(v_l, Y_j) / d(v_l, Y_j)]^2},
$$

if $d(v_i, Y_j) > 0$ for all $i = 1, 2, \ldots, k$ and $\mu_{ij}^{(i)} = 0$ for otherwise ($d(v_i, Y_j)$ is the Euclidean distance from $v_i$ to $Y_j$).

Step 3.4. Repeat Step 3.2 and Step 3.3 until the following condition is satisfied:

$$
\max_{j} \left( \left| \mu_{ij}^{(i+1)} - \mu_{ij}^{(i)} \right| \right) < \varepsilon.
$$

When Step 3 stops, we will be obtained a matrix of size $(k \times n)$. In this matrix, we have the sum of each column that always equal 1 ($\sum \mu_{ij} = 1$). If $\max \{ \mu_{ij} \} = \mu_{im}, 1 \leq m \leq k$, then the element $Y_{ij}, 1 \leq i \leq n$ is put in cluster $C_m, 1 \leq m \leq k$.

Step 4. Utilize $c_i = v_i, i = 1, 2, \ldots, k$ from step 3 and determine the universal set $U$ as follows:

$$
U = [c_1 - \frac{c_2 - c_1}{2}, c_2 + \frac{c_2 - c_1}{2}, \ldots, c_k - \frac{c_k - c_{k-1}}{2}, c_k + \frac{c_k - c_{k-1}}{2}].
$$

Divide the universal set $U$ into $k$ intervals:

$$
U_i = [c_1 - \frac{c_2 - c_1}{2}, c_2 + \frac{c_2 - c_1}{2}] = [D_1; D_2]; U_i = [c_{i-1} + c_i, c_i + c_{i+1} + c_i] = [D_i; D_{i+1}], i = 2, k - 1
$$

$$
U_k = [c_{k-1} + c_k, c_k + \frac{c_k - c_{k-1}}{2}] = [D_{k-1}; D_k]
$$

Step 5. Calculate the different mean of intervals by the following equation:

$$
\bar{D} = \frac{1}{k} \sum_{i=1}^{k} (D_i - D_{i+1}).
$$

With $k$ intervals ($k$-fuzzy numbers) and $A_1, A_2, \ldots, A_k$ can be defined as follows:

$$
A_1 = [D_1 - \bar{D}, D_1, D_2, D_2 + \bar{D}];
$$

$$
A_2 = [D_2 - \bar{D}, D_2, D_3, D_3 + \bar{D}];
$$

$$
\ldots
$$

$$
A_k = [D_{k-1} - \bar{D}, D_{k-1}, D_k, D_k + \bar{D}];
$$

where $A_i, i = 1, k$ is the fuzzy numbers.

Step 6. If $Y_j \in U_i, 1 \leq i \leq k, 1 \leq j \leq N$ then $Y_j \rightarrow U_i$ and the fuzzy logical relationship of $k$ intervals is $U_i(Y_j < Y_i) \rightarrow U_i$.

- If the fuzzy logical relationship group of $U_i$ is empty ($U_i \rightarrow \emptyset$) then $FI = A_i$.
- If the fuzzy logical relationship group of $U_i$ is one-to-one then $FI = A_{i-1}$.
- If the fuzzy logical relationship group of $U_i$ is one-to-many,

$$
U_i \rightarrow U_i, U_i \rightarrow U_a, \ldots, U_i \rightarrow U_b, 1 \leq i, a, b \leq k
$$

then $FI = (A_i + A_a + A_b) / (i + a + b)$. 


Step 7. Calculate the forecasting interval for time series by the following equation

\[ FY = FI \times \frac{\max(X_i)}{100}, \]

where, \( FI = [FI_{1,j}, FI_{2,j}] \) and

\[ FI_{1,j} = (D_j - D_{j-1}) \times \alpha + D_{j-1}, \]

\[ FI_{2,j} = D_k - (D_k - D_{k-1}) \times \alpha, \]

\( \alpha \) is the degree of confidence of forecasted intervals.

4. The numerical example

In this section, we use the Enrollment data of Alabama University (ACD) to illustrate the proposed algorithm. This data is used in many researches as Song and Chisson [18], Chen [8], Huarng [13], Singh [17] and it is shown in the Table 1.

| Year | X_i  | Y_i   | V(18) | Cluster | Year | X_i  | Y_i   | V(18) | Cluster |
|------|------|-------|-------|---------|------|------|-------|-------|---------|
| 1971 | 13055| 67.513| 67.513| C_1     | 1982 | 15433| 79.811| 79.964| C_5     |
| 1972 | 13563| 70.140| 70.140| C_2     | 1983 | 15497| 80.142| 79.964| C_5     |
| 1973 | 13867| 71.712| 71.712| C_3     | 1984 | 15145| 78.321| 78.372| C_5     |
| 1974 | 14696| 75.999| 75.999| C_4     | 1985 | 15163| 78.414| 78.372| C_11    |
| 1975 | 15460| 79.950| 79.964| C_5     | 1986 | 15984| 82.660| 82.621| C_12    |
| 1976 | 15311| 79.180| 79.198| C_6     | 1987 | 16859| 87.185| 87.181| C_9     |
| 1977 | 15603| 80.690| 80.631| C_7     | 1988 | 18150| 93.862| 93.862| C_13    |
| 1978 | 15861| 82.024| 82.064| C_8     | 1989 | 18970| 98.102| 97.859| C_14    |
| 1979 | 16807| 86.916| 87.181| C_9     | 1990 | 19328| 99.953| 99.977| C_15    |
| 1980 | 16919| 87.495| 87.181| C_9     | 1991 | 19337| 100.000| 99.977| C_15    |
| 1981 | 16388| 84.749| 84.749| C_10    | 1992 | 18876| 97.616| 97.859| C_14    |
| 1972 | 13563| 70.140| 70.140| C_2     |

The steps of the proposed algorithm are presented as follows:

Step 1: Standardize the time series on the scale 100, we have the result presented at column Y_i of Table 1.

Step 2: Determine the number of clusters for time series by the SNC algorithm. This result is shown by Figure 1.

Figure 1 shows that Enrollment data converge into 15 clusters. The numeral result obtained in Table 1.

Step 3: For 15 clusters obtained in Step 2, we determine the centroid points of each cluster according to the SNC algorithm. The outcome is shown in Table 2.
**Table 2. The value of centroid clusters**

| Cluster | Centroid ($v_i$) | Cluster | Centroid ($v_i$) |
|---------|------------------|---------|------------------|
| $C_1$   | 67.513           | $C_9$   | 87.181           |
| $C_2$   | 70.140           | $C_{10}$| 84.749           |
| $C_3$   | 71.712           | $C_{11}$| 78.372           |
| $C_4$   | 75.999           | $C_{12}$| 82.621           |
| $C_5$   | 79.964           | $C_{13}$| 93.862           |
| $C_6$   | 79.198           | $C_{14}$| 97.859           |
| $C_7$   | 80.631           | $C_{15}$| 99.977           |
| $C_8$   | 82.064           |         |                  |

Table 2 shows the value of 15 representing clusters for time series.

**Step 4.** Divide the universal set $U$ into 15 intervals as follows:

$U_1$ = [66.1995; 68.8266], $U_2$ = [68.8266; 70.9262], $U_3$ = [70.9262; 73.8558], $U_4$ = [73.8558; 77.9819], $U_5$ = [77.9819; 79.5810], $U_6$ = [79.5810; 79.9142], $U_7$ = [79.9142; 81.3472], $U_8$ = [81.3472; 84.6222], $U_9$ = [84.6222; 85.9651], $U_{10}$ = [85.9651; 81.5605], $U_{11}$ = [81.5605; 83.6850], $U_{12}$ = [83.6850; 85.9651], $U_{13}$ = [85.9651; 88.2410], $U_{14}$ = [88.2410; 95.8604], $U_{15}$ = [95.8604; 98.9180]

**Step 5.** Calculate $D = 2.3224$ and $A_i$, $i = 1, ..., 15$, we have the Table 3:

**Table 3. The computing result of $D$**

| $A_1$  | 63.8771 | 66.1995 | 68.8266 | 71.1490 |
|--------|---------|---------|---------|---------|
| $A_2$  | 66.5042 | 68.8266 | 70.9262 | 73.2486 |
| $A_3$  | 68.6038 | 70.9262 | 73.8558 | 76.1782 |
| $A_4$  | 71.5334 | 73.8558 | 77.1855 | 79.5079 |
| $A_5$  | 74.8631 | 77.1855 | 78.7846 | 81.1070 |
| $A_6$  | 76.4622 | 78.7846 | 79.5810 | 81.9034 |
| $A_7$  | 77.2586 | 79.5810 | 80.2976 | 82.6200 |
| $A_8$  | 77.9752 | 80.2976 | 81.3472 | 83.6696 |
| $A_9$  | 79.0249 | 81.3472 | 82.3421 | 84.6645 |
| $A_{10}$| 80.0197 | 82.3421 | 83.6850 | 86.0074 |
| $A_{11}$| 81.3626 | 83.6850 | 85.9651 | 88.2875 |
| $A_{12}$| 83.6427 | 85.9651 | 90.5211 | 92.8435 |
| $A_{13}$| 88.1987 | 90.5211 | 95.8604 | 98.1828 |
| $A_{14}$| 93.5380 | 95.8604 | 98.9180 | 101.2404 |
| $A_{15}$| 96.5956 | 98.9180 | 101.0354 | 103.3578 |

**Step 6.** Establish the fuzzy logical relationship as follows:
Table 4. The fuzzy logical relationship of elements

| Year | Interval 1 | Interval 2 |
|------|------------|------------|
| 1971 | [63.877, 66.200] | [68.827, 71.149] |
| 1972 | [66.504, 68.827] | [70.926, 73.249] |
| 1973 | [68.604, 70.926] | [73.856, 76.178] |
| 1974 | [71.533, 73.856] | [77.186, 79.508] |
| 1975 | [77.259, 79.581] | [83.643, 85.965] |
| 1976 | [77.219, 79.541] | [80.464, 82.787] |
| 1977 | [77.975, 80.298] | [81.347, 83.670] |
| 1978 | [76.944, 79.266] | [80.563, 82.886] |
| 1979 | [83.643, 85.965] | [90.521, 92.844] |
| 1980 | [84.401, 86.724] | [90.782, 93.105] |

Find solution of the fuzzy logical relationship.

Table 5. The result of fuzzy numbers

| Fuzzy numbers | Interval 1 | Interval 2 |
|---------------|------------|------------|
| $U_1 \rightarrow U_2$ | [63.877, 66.200] | [68.827, 71.149] |
| $U_2 \rightarrow U_3$ | [66.504, 68.827] | [70.926, 73.249] |
| $U_3 \rightarrow U_4$ | [68.604, 70.926] | [73.856, 76.178] |
| $U_4 \rightarrow U_5$ | [71.533, 73.856] | [77.186, 79.508] |
| $U_5 \rightarrow U_6$ | [77.259, 79.581] | [83.643, 85.965] |
| $U_6 \rightarrow U_7$ | [77.219, 79.541] | [80.464, 82.787] |
| $U_7 \rightarrow U_8$ | [77.975, 80.298] | [81.347, 83.670] |
| $U_8 \rightarrow U_9$ | [76.944, 79.266] | [80.563, 82.886] |
| $U_9 \rightarrow U_{10}$ | [83.643, 85.965] | [90.521, 92.844] |
| $U_{10} \rightarrow U_{11}$ | [90.782, 93.105] | [98.918, 101.240] |

Calculate the value of the FI

Table 6. The outcome of the FI

| Year | Interval 1 | Interval 2 |
|------|------------|------------|
| 1971 | [63.877, 66.200] | [68.827, 71.149] |
| 1972 | [66.504, 68.827] | [70.926, 73.249] |
| 1973 | [68.604, 70.926] | [73.856, 76.178] |
| 1974 | [71.533, 73.856] | [77.186, 79.508] |
| 1975 | [77.259, 79.581] | [83.643, 85.965] |
| 1976 | [77.219, 79.541] | [80.464, 82.787] |
| 1977 | [77.975, 80.298] | [81.347, 83.670] |
| 1978 | [76.944, 79.266] | [80.563, 82.886] |
| 1979 | [83.643, 85.965] | [90.521, 92.844] |
| 1980 | [84.401, 86.724] | [90.782, 93.105] |
Step 7. The time series is predicted by the intervals with 80% of the degree of confidence that is shown in Table 7 below.

Table 7. The outcome of forecasting according to interval

| Year | FY       | Year   | FY       |
|------|----------|--------|----------|
| 1971 | [12711.183, 13398.817] | 1982   | [15298.755, 15616.957] |
| 1972 | [13219.185, 13804.816] | 1983   | [15291.044, 15649.161] |
| 1973 | [13625.184, 14371.315] | 1984   | [15237.922, 15668.353] |
| 1974 | [14191.638, 15015.177] | 1985   | [15334.111, 15798.189] |
| 1975 | [15298.755, 15616.957] | 1986   | [15334.111, 15798.189] |
| 1976 | [15291.044, 15649.161] | 1987   | [16533.252, 17593.884] |
| 1977 | [15437.324, 15819.934] | 1988   | [16679.951, 17644.370] |
| 1978 | [15237.922, 15668.353] | 1989   | [18446.710, 19217.593] |
| 1979 | [16533.252, 17593.884] | 1990   | [19037.960, 19627.040] |
| 1980 | [16679.951, 17644.370] | 1991   | [18742.335, 19422.316] |
| 1981 | [16679.951, 17644.370] |        |          |

Figure 2. The line graph presents the comparison between actual outcome and upper and lower bound of forecasting result.

Figure 2 shows time series predicted based on the confidence level is 80%. Furthermore, they also present the fluctuating level of original data according to upper and lower bound.

Compare to the forecasting result of some method, we have the Table 8:

Table 8. The result of algorithms for the Enrollment dataset

| Criteria | Chen (1996) [8] | Huang (2001) [13] | Abbasov (2003) [1] | Lee (2004) [15] | Chen (2004) [7] |
|----------|----------------|------------------|-------------------|----------------|----------------|
| MAE      | 502.38         | 299.15           | 479.57            | 296.15         | 293.45         |
| MAPE     | 3.08           | 2.45             | 2.87              | 2.69           | 1.76           |
| MSE      | 413,980.98     | 226,611          | 342,326           | 255,227        | 138,366.80     |

| Criteria | Singh (2007) [17] | Yu (2010) [23] | Khashei (2011) [14] | Chen (2013) [9] | Ghosh (2015) [12] |
|----------|------------------|---------------|-------------------|----------------|------------------|
| MAE      | 254.16           | 216.50        | 211.12            | 314.34         | 298.68           |
| MAPE     | 1.53             | 2.15          | 2.12              | 2.17           | 1.82             |
| MSE      | 95,305           | 47,231.03     | 31,021            | 41,235         | 186,421          |
According to Table 6, we can see that the proposed model has the better outcome than others because it has the lowest parameters.

5. Conclusion
The article has presented a fuzzy forecasting model for time series, a common data type in reality. Based on many important improvements from the unsupervised learning algorithm, the existing fuzzy time series, and the trapezoidal fuzzy numbers, we propose the new interval forecasting model. In addition, this study has applied well machine learning technique in training for fuzzy time series. This is a new research direction in the current digital age. The proposed model has performed better than the existing ones based on the Enrollment dataset. With this new model, in the next time, we will continue to improve the steps to divide the domain time series and the training data based on improving the machine learning technique, and apply to many important forecasts to reality.

Acknowledgments
For Dinh PhamToan, this work was funded by Vingroup Joint Stock Company and supported by the Domestic Master/ PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), Vingroup Big Data Institute (VINBIGDATA)

References
[1] Abbasov A and Mamedova M 2003 Application of fuzzy time series to population forecasting. Vienna University of Technology 1 pp 545-552
[2] Abreu P H, Silva D C, Mendes-Moreira J, Reis L P and Garganta J 2013 Using multivariate adaptive regression splines in the construction of simulated soccer team’s behavior models. International Journal of Computational Intelligence Systems 6 pp 893-910
[3] Aladag S, Aladag C H, Mentes T and Egrioglu E 2012 A new seasonal fuzzy time series method based on the multiplicative neuron model and SARIMA, Hacettepe Journal of Mathematics and Statistics 41 pp 145-163
[4] Bindu G, Rohit G 2016 Enhanced accuracy of fuzzy time series model using ordered weighted aggregation. Appl. Soft Comput 48 pp 265-280
[5] Box G E P and Jenkins G M 1970 Time series analysis: Forecasting and control, Holden-Day, San Francisco 546 pages
[6] Chen J and Hung W 2015 An automatic clustering algorithm for probability density functions, J. Stat. Comput. Simul 48 pp 3047-3063
[7] Chen S M and Hsu C C 2004 A new method to forecast enrollments using fuzzy time series, International Journal of Applied Science and Engineering 2 pp 234–244
[8] Chen S M 1996 Forecasting enrollments based on fuzzy time series, Fuzzy sets and systems 81 pp 311–319
[9] Chen S M, Kao P Y 2013 TAIEX forecasting based on fuzzy time series, Particle swarm optimization techniques and support vector machines, Inf. Sci. 247 pp 62-71
[10] Egrioglu S, Bas E, Aladag C H, Yolcu U 2016 Probabilistic fuzzy time series method based on artificial neural network, American Journal of Intelligent Systems 62 pp 42-47
[11] Eren B, Vedide R, Uslu U and Erol E 2014 A modified genetic algorithm for forecasting fuzzy time series, Applied Intelligence 41 pp 453–463
[12] Ghosh H, Chowdhury S, Prajneshu S 2015 An improved fuzzy time-series method of forecasting
based on L-R fuzzy sets and its application, *Journal of Applied Statistics* 43 pp 1128–1139

[13] Huarng K 2001 Heuristic models of fuzzy time series for forecasting, *Fuzzy sets and systems* 123 pp 369–386

[14] Khashei M, Bijari M, Hejazi C S R 2011 An extended fuzzy artificial neural networks model for time series forecasting, *Iranian Journal of Fuzzy Systems* 3 pp 45-66

[15] Lee H S, Chou M T 2004 Fuzzy forecasting based on fuzzy time series, *International Journal of Computer Mathematics* 81 pp 781-789

[16] Liu Hao-Tien 2007 An improved fuzzy time series forecasting method using trapezoidal fuzzy numbers, *Fuzzy Optimization and Decision Making* 6 pp 63-80

[17] Singh S R 2007 A simple method of forecasting based on fuzzy time series, *Applied mathematics and computation* 186 pp 330–339

[18] Song Q and Chissom B S 1993 Fuzzy time series and its models, *Fuzzy sets and systems* 54 pp 269-277

[19] Tai Vovan 2018 An improved fuzzy time series forecasting model using variations of data, *Fuzzy Optimization and Decision Making* 18 pp 151-173

[20] Trendafilova I and Manoach E 2008 Vibration-based damage detection in plates by using time series analysis, *Mechanical Systems and Signal Processing* 22 pp 1092-1106

[21] Tseng F M, Tzeng G H and Yu H C 1999 Fuzzy seasonal time series for forecasting the production value of the mechanical industry in Taiwan, *Technological Forecasting and Social Change* 60 pp 263-273

[22] Trendafilova I 2006 Vibration-based damage detection in structures using time series analysis, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 220 pp 261-272

[23] Yu H K, Huarng K 2010 A neural network-based fuzzy time series model to improve forecasting, *Expert Systems with Applications* 37 pp 3366-3372

[24] Yusuf S M, Mohammad A, Hamisu A A 2017 A novel two-factor high order fuzzy time series with applications to temperature and futures exchange forecasting, *Nigerian Journal of Technology* 36 pp 1124-1134