A New Method For Forecasting Enrolments Combining Time-Variant Fuzzy Logical Relationship Groups And K-Means Clustering

Nghiem Van Tinh[1], Vu Viet Vu[1], Tran Thi Ngoc Linh[1]

[1] *Thai Nguyen University of Technology, Thai Nguyen University*

Abstract – *In this paper, a new forecasting model based on two computational methods, time-variant fuzzy logical relationship groups and K-mean clustering technique, is presented for academic enrolments. Firstly, we use the K-mean clustering algorithm to divide the historical data into clusters and adjust them into intervals with different lengths. Then, based on the new intervals, we fuzzify all the historical data of the enrolments of the University of Alabama and calculate the forecasted output by the proposed method. Compared to the other methods existing in literature, particularly to the first-order fuzzy time series, the proposed method showed a better accuracy in forecasting the number of students in enrolments of the University of Alabama from 1971s to 1992s.*

Key Words: *Fuzzy time series, Fuzzy forecasting, Fuzzy logic relationship, K-means clustering, enrolments*

1. INTRODUCTION

In the past decades, many forecasting models have been developed to deal with various problems in order to help people to make decisions, such as crop forecast [7], [8] academic enrolments [2], [11], the temperature prediction [14], stock markets[15], etc. There is the matter of fact that the traditional forecasting methods cannot deal with the forecasting problems in which the historical data are represented by linguistic values. Ref. [2,3] proposed the time-invariant fuzzy time and the time-variant time series model which use the max–min operations to forecast the enrolments of the University of Alabama. However, the main drawback of these methods is huge computation burden. Then, Ref. [4] proposed the first-order fuzzy time series model by introducing a more efficient arithmetic method. After that, fuzzy time series has been widely studied to improve the accuracy of forecasting in many applications. Ref. [5] considered the trend of the enrolment in the past years and presented another forecasting model based on the first-order fuzzy time series. At the same time, Ref. [9],[12] proposed several forecast models based on the high-order fuzzy time series to deal with the enrolments forecasting problem. In [10], the length of intervals for the fuzzy time series model was adjusted to get a better forecasted accuracy. Ref.[13] presented a new forecast model based on the trapezoidal fuzzy numbers. Ref.[19] shown that different lengths of intervals may affect the accuracy of forecast. Recently, Ref.[17] presented a new hybrid forecasting model which combined particle swarm optimization with fuzzy time series to find proper length of each interval. Additionally, Ref.[18] proposed a new method to forecast enrollments based on automatic clustering techniques and fuzzy logical relationships.

In this paper, we proposed a new forecasting model combining the time-variant fuzzy relationship groups and K-mean clustering technique. The method is different from the approach in [4] and [17] in the way where the fuzzy relationships are created. Based on the model proposed in [10], we have developed a new weighted fuzzy time series model by combining the clustering technique K-mean and time-variant fuzzy relationship groups with the aim to increase the accuracy of the forecasting model.

In case study, we applied the proposed method to forecast the enrolments of the University of Alabama. The experimental results show that the proposed method gets a higher average forecasting accuracy compared to the existing methods.

The remainder of this paper is organized as follows. In Section 2, we provide a brief review of fuzzy time series and K-means clustering technique. In Section 3, we present our method for forecasting the enrolments of the University of Alabama based on the K-means clustering algorithm and time-variant fuzzy logical relationship groups. Then, the experimental results are shown and analyzed in Section 4. Conclusions are presented in Section 5

2. FUZZY TIME SERIES AND K-MEANS CLUSTERING

2.1 Fuzzy time serses

Fuzzy set theory was firstly developed by Zadeh in the 1965s to deal with uncertainty using linguistic terms. Ref.[2] successfully modelled the fuzzy forecast by adopting the fuzzy sets for fuzzy time series. To avoid complicated max–min composition operations, in[4] improved the fuzzy forecasting method by using simple arithmetic operations. Let $U=\{u_1,u_2,\ldots,u_n\}$ be an universal set; a fuzzy set A of U is defined as $A=\{f_A(u_1)/u_1+\ldots+f_A(u_n)/u_n\}$, where f_A is a membership function of a given set A, $f_A: U\rightarrow[0,1]$, $f_A(u_i)$ indicates the grade of membership of u_i in the fuzzy set A, $f_A(u_i) \in [0, 1]$, and $1\le i \le n$. General definitions of fuzzy time series are given as follows:

Definition 2.1: Fuzzy time series

Let $Y(t)$ $(t = .., 0, 1, 2 ..)$, a subset of R, be the universe of discourse on which fuzzy sets $f_i(t)$ (i = 1,2…) are defined and if $F(t)$ be a collection of $f_i(t)$) (i = 1, 2…). Then, $F(t)$ is called a fuzzy time series on $Y(t)$ (t . . . . . 0, 1,2 . . . .).

Definition 2.2: Fuzzy logic relationship

If there exists a fuzzy relationship R(t-1,t), such that F(t) = F(t-1)∗R(t-1,t), where "∗" is an arithmetic operator, then F(t) is said to be caused by F(t-1). The relationship between F(t) and F(t-1) can be denoted by F(t-1)→ F(t). Let A_i = F(t) and A_j = F(t-1), the relationship between F(t) and F(t -1) is denoted by fuzzy logical relationship $A_i \rightarrow$ Aj where A_i and A_j refer to the current state or the left hand side and the next state or the right-hand side of fuzzy time series.

Definition 2.3: λ- Order Fuzzy Relations

Let *F(t)* be a fuzzy time series. If F(t) is caused by F(t-1), F(t-2),..., F(t-λ+1) F(t-λ) then this fuzzy relationship is represented by by F(t-λ), ..., F(t-2), F(t-1)→ F(t) and is called an λ- order fuzzy time series.

Definition 2.4: Time-Invariant Fuzzy Time Series

Let *F(t)* be a fuzzy time series. If for any time *t, F(t) = F(t* - 1) and *F(t)* only has finite elements, then *F(t)* is called a *time-invariant* fuzzy time series. Otherwise, it is called a *time-variant* fuzzy time series.

Definition 2.5: Fuzzy Relationship Group (FLRG)

Fuzzy logical relationships, which have the same left-hand sides, can be grouped together into fuzzy logical relationship groups. Suppose there are relationships such that

$A_i \rightarrow A_k$

$$A_i \rightarrow A_m$$

....

So, base on [4], these fuzzy logical relationship can be grouped into the same FLRG as : $A_i \rightarrow A_k, A_m$ ...

Definition 2.6: Time-variant fuzzy relationship groups

The fuzzy relationship is determined by the relationship of F(t-1)→ $F(t)$. If, let $F(t) = A_i(t)$ and $F(t-1) = A_j(t-1)$, we will have the relationship $A_j(t-1) \rightarrow A_i(t)$. At the time t, we have the following fuzzy relationship: $A_j(t-1) \rightarrow A_i(t),\ A_j(t1-1) \rightarrow A_{i1}(t1), \ldots, A_j(tp-1) \rightarrow A_{ip}(tp)$ with $t1, t2, \ldots, tp \leq t$. It means that if the fuzzy relationship took place before $A_j(t-1) \rightarrow A_i(t)$, we can group the fuzzy logic relationship to be $A_j(t-1) \rightarrow A_i(t), A_{i1}(t1), A_{i2}(t2), A_{ip}(tp)$. It is called time-variant fuzzy logic relationship group.

Definition 2.7: Forecast Error

The forecasting error is the difference between the actual value and the prediction in time series.

Error = Actual value - forecasted value

2.2 K-Means clustering technique

K-means clustering introduced in [1] is one of the simplest unsupervised learning algorithms for solving the well-known clustering problem. K-means clustering method groups the data based on their closeness to each other according to Euclidean distance. In this clustering approach, the user decides that how many clusters should be and on the basis of closeness of the data vector to the centroid, which is the mean of the data vector of clusters is assigned to its own which shows the minimum distance. The result depends on the number of cluster (J value) and the initial centroid chosen by the K-means algorithm

- The algorithm is composed of the following steps

  Step 1: Place j points into the space represented by the objects that are being clustered. These points represent initial group centroids.

  Step 2: Assign each object to the group that has the closest centroid.

  Step 3: When all objects have been assigned, recalculate the positions of the j centroids.

  Step 4: Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

3. FUZZY TIME SERIES MODEL BASE ON K-MEANS CLUSTERING

In this section, we presents a new method for forecasting the enrolments of University of Alabama based on time series fuzzy relationship groups and K-means clustering techniques. At first, we apply K-means clustering technique to classify the collected data into clusters and adjust these clusters into contiguous intervals for generating intervals from numerical data then, based on the interval defined, we fuzzify on the historical data determine fuzzy relationship and cluster time-variant fuzzy relationship; and finally, we obtain the forecasting output based on the time-variant fuzzy relationship groups and rules of forecasting are our propsed.The historical data of enrollments of the University of Alabama are listed in Table 1.

Table 1: Historical data of enrollments of the University of Alabama

Year	Actual Enrollment
1971	13055
1972	13563
1973	13867
1974	14696
1975	15460
1976	15311
1977	15603
1978	15861
1979	16807
1980	16919
1981	16388
1982	15433
1983	15497
1984	15145
1985	15163
1986	15984
1987	16859
1988	18150
1989	18970
1990	19328

1991	19337
1992	18876

Source: In [4], [7] and [2-3]

3.1 The K-Mean clustering algorithm for generating intervals from historical data

The algorithm composed of 4 steps is introduced step-by-step with the same dataset.

Step 1: Apply the K-means clustering algorithm to partition the historical time series data into p clusters and sort the data in clusters in an ascending sequence. in this paper, we set p=14 clusters, the results are as follows: {13055,13563}, {13867}, {14696}, {15145,15163}, {15460, 15311,15433,15497}, {15603}, {15861,15984}, {16388}, {16807}, {16859}, {16919}, {18150}, {18970,18876}, {19328, 19337}

Step 2: Create the cluster center

In this step, we use automatic clustering techniques [18] to generate cluster center($Center_j$) from clusters in step 1 shown in Table 2 of each cluster $cluster_j$ as follows:

$$Center_j = \frac{\sum_{i=1}^{n} d_i}{n} \quad (2)$$

where d_i is a datum in $Cluster_j$, n denotes the number of data in $Cluster_j$ and $1 \leq j \leq p$.

Step 3: Adjust the clusters into intervals according to the follow rules. Assume that $Center_k$ and $Center_{k+1}$ are adjacent cluster centers, then the upper bound $cluster\_UB_j$ of $cluster_j$ and the lower bound $cluster\_LB_{k+1}$ of $cluster_{j+1}$ can be calculated as follows:

$$Cluster\_UB_k = \frac{Center_k + Center_{k+1}}{2} \quad (3)$$

$$Cluster\_LB_{k+1} = Cluster\_UB_k \quad (4)$$

where k = 1,.., p-1. Because there is no previous cluster before the first cluster and there is no next cluster after the last cluster, the lower bound $Cluster\_LB_1$ of the first cluster and the upper bound $Cluster\_UB_p$ of the last cluster can be calculated as follows:

$$Cluste\_LB_1 = Center_1 - (Center_1 - Cluster\_UB_1) \quad (5)$$

$$Cluste\_UB_p = Center_p + (Center_p - Cluster\_LB_p) \quad (6)$$

Table 2: Generate cluster center from clusters

No	Group	Center
1	{13055, 13563}	13309
2	{13867}	13867
3	{14696}	14696
4	{15145, 15163}	15154
5	{15460, 15311, 15433, 15497 }	15425
6	{15603 }	15603
7	{15861, 15984 }	15922
8	{16388}	16388
9	{16807}	16807
10	{16859}	16859
11	{16919}	16919
12	{18150 }	18150
13	{18970, 18876 }	18923
14	{19328, 19337 }	19332

Step 4: Let each cluster $Cluster_j$ form an interval $interval_j$, which means that the upper bound $Cluster\_UB_j$ and the lower bound $Cluster\_LB_j$ of the cluster $cluster_j$ are also the upper bound $interval\_UBound_j$ and the lower bound $interval\_LBound_j$ of the interval $interval_j$, respectively. Calculate the middle value Mid_value_j of the interval $interval_j$ as follows:

$$Mid\_value_j = \frac{interval\_LB_j + interval\_UB_j}{2} \quad (7)$$

where $interval\_UB_j$ and $interval\_LB_j$ are the upper bound and the lower bound of $interval_j$, respectively, and j = 1,.. ,p.

Table 3: The midpoint of each interval $u_i (1 \leq j \leq 14)$

No	Intervals	M_j = Mid_value
1	(13030, 13588]	13309
2	(13588, 14282]	13935
3	(14282, 14925]	14603.5
4	(14925, 15290]	15107.5
5	(15290, 15514]	15402
6	(15514, 15762]	15638
7	(15762, 16155]	15958.5
8	(16155, 16598]	16376.5
9	(16598, 16833]	16715.5
10	(16833, 16889]	16861
11	(16889, 17534]	17211.5
12	(17534, 18536]	18035
13	(18536, 19128]	18832
14	(19128, 19536]	19332

3.2 Forecasting enrollments using the proposed method

In this section, we present a new method for forecasting enrollments based on the K-mean clustering algorithm and time variant fuzzy relationships group. The proposed method is now presented as follows:

Step 1: *Partition the universe of discourse into intervals*

After applying the procedure K-mean clustering, we can get the following 14 intervals and calculate the middle value of the intervals are listed in Table 3.

Step 2: Fuzzify all historical data.

Define each fuzzy set A_i based on the new obtained 14 intervals in step 1 and the historical enrollments shown in Table 1. For 14 intervals, there are 14 linguistic variables A_i. Each linguistic variable represents a fuzzy set such that the according to Eq(8). additionally, we use a triangular function to define the fuzzy sets A_i. It can be illustrated in Fig -1 and Each historical value is fuzzified according to its highest degree of membership. If the highest degree of belongingness of a certain historical time variable, say $F(t-1)$ occurs at fuzzy set A_i , then $F(t-1)$ is fuzzified as A_i

$$A_i = \sum_{j=1}^{m} \frac{a_{ij}}{u_j} = \begin{cases} 1 & \text{if } j == i \\ 0.5 & \text{if } j == i-1 \text{ or } j == i+1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where $a_{ij} \in [0,1]$, $1 \leq i \leq m$, and $1 \leq j \leq m$, $m=14$. The value of a_{ij} indicates the grade of membership of u_j in the fuzzy set A_i. For example, the historical enrollment of year 1973 is 13867 which falls within u_2 = (13588, 14282], so it belongs to interval u_2 Based on Eq. (8), Since the highest membership degree of u_2 occurs at A_2, the historical time variable $F(1973)$ is fuzzified as A_2. The results of fuzzification are listed in Table 4, where all historical data are fuzzified to be fuzzy sets.

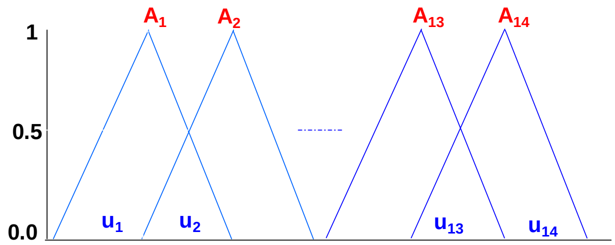

Fig -1: Membership function for the enrollments

Table 4: Linguistic values for the enrollments of the University of Alabama

Year	Actual Enrollment	Linguistic values
1971	13055	A1
1972	13563	A1
1973	13867	A2
1974	14696	A3
1975	15460	A5
1976	15311	A5
1977	15603	A6
1978	15861	A7
1979	16807	A9
1980	16919	A11
1981	16388	A8
1982	15433	A5
1983	15497	A5
1984	15145	A4
1985	15163	A4
1986	15984	A7
1987	16859	A10
1988	18150	A12
1989	18970	A13
1990	19328	A14
1991	19337	A14
1992	18876	A13

Let Y(t) be a historical data time series on year t. The purpose of Step 1 is to get a fuzzy time series F(t) on Y(t). Each element of Y(t) is an integer with respect to the actual enrollment. But each element of F(t) is a linguistic value (i.e. a fuzzy set) with respect to the corresponding element of Y(t). For example, in Table 4, Y(1971) = 13055 and F(1971) = A_1; Y(1973) = 13867 and F(1973) = A_2; Y(1975) = 15460 and F(1975) = A_5 and so on.

Step 3: Create all fuzzy relationships

Relationships are identified from the fuzzified historical data. So, from Table 4 and base on Definition 2.2, we get first – order fuzzy logical relationships are shown in Table 5. where the fuzzy logical relationship $A_i \rightarrow A_k$ means "If the enrollment of year i is A_i, then that of year i + 1 is A_k", where A_i is called the current state of the enrollment, and A_k is called the next state of the enrollment (Note that even though the same relationships may appear more than once).

Table 5: The first-order fuzzy logical relationships on the enrollments

No	Relationships
1	A1 -> A1
2	A1 -> A2
3	A2 -> A3
4	A3 -> A5
5	A5 -> A5
6	A5 -> A6
7	A6 -> A7
8	A7 -> A9
9	A9 -> A11
10	A11 -> A8
11	A8 -> A5
12	A5 -> A5
13	A5 -> A4
14	A4 -> A4
15	A4 -> A7
16	A7 -> A10
17	A10 -> A12
18	A12 -> A13
19	A13 -> A14
20	A14 -> A14
21	A14 -> A13

Step 4: Establish all fuzzy logical relationship groups

By Chen [4], all the fuzzy relationship having the same fuzzy set on the left-hand side or the same current state can be put together into one fuzzy relationship group. But, according to the Definition 2.6, we need to consider the appearance history of the fuzzy sets on the right-hand side too. Therefore, only the element on the right hand side appearing before the left-hand side of the relationship group is taken into the same fuzzy logic relationship group.Thus, from Table 5 and based on Definition 2.6 (Time-variant fuzzy relationship group), we can obtain 21 fuzzy logical relationship groups shown in Table 6.

Table 6: Fuzzy logical relationship groups

Groups	Fuzzy relationship groups
G1	A1
2	A1, A2
3	A3
4	A5
5	A5
6	A5, A6
7	A7
8	A9
9	A11
10	A8
11	A5
12	A5, A6, A5
13	A5, A6, A5, A4
14	A4
15	A4, A7
16	A9, A10
17	A12
18	A13
19	A14
20	A14
21	A14, A13

Step 5: Calculate the forecasting output.

Calculate the forecasted output at time t by using the following principles:

Principle 1: If the fuzzified enrollment of year t-1 is A_j and there is only one fuzzy logical relationship in the fuzzy logical relationship group whose current state is A_j, shown as follows: $A_j(t-1) \rightarrow A_k(t)$

then the forecasted enrollment of year t is m_k, where m_k is the midpoint of the interval u_k and the maximum membership value of the fuzzy set A_k occurs at the interval u_k

Principle 2: If the fuzzified enrollment of year t -1 is A_j and there are the following fuzzy logical relationship group whose current state is A_j, shown as follows:

$A_j(t-1) \rightarrow A_{i1}(t1), A_{i2}(t2), A_{ip}(tk)$

then the forecasted enrollment of year t is calculated as follows: $forecasted = \frac{1*m_{i1}+2*m_{i2}+3*m_{i3}+\cdots+p*m_{ip}}{1+2+\cdots+p}$;

where m_{i1}, m_{i2}, m_{ik} are the middle values of the intervals u_{i1}, u_{i2} and u_{ip} respectively, and the maximum membership values of $A_{i1}, A_{i2}, \ldots, A_{ip}$ occur at intervals u_{i1}, u_{i2}, u_{ip}, respectively.

Principle 3: If the fuzzified enrollment of year t-1 is A_j and there is a fuzzy logical relationship in the fuzzy logical relationship group whose current state is A_j, shown as follows: $A_j(t-1) \rightarrow \neq$

then the forecasted enrollment of year t is m_j, where m_j is the midpoint of the interval u_j and the maximum membership value of the fuzzy set A_j occurs at the interval u_j

Based on Tables 4 and 6, we can forecast the enrollments of the University of Alabama from 1971 to 1992 by the proposed method: For example, the forecasted enrollments of the years 1973, 1975 and 1981 are calculated as follows:

[1973] From Table 4, we can see that the fuzzified enrollments of years F(t-1)= F(1972) is A_1 . From Table 6, we can see that there is a fuzzy logical relationship $A_1(t-1) \rightarrow A_1(t), A_2(t)$, in Group 2. Based on Principle 2, the forecasted enrollment of year 1973 can be calculated as follows:

$$Forecasted = \frac{1*m_1+2*m_2}{1+2} = \frac{1*13309+2*13935}{3} = 13726.33$$

Moreover, assume that we want to forecast the enrollment of year 1977, then from Table 4, we can see that the fuzzified enrollment of year F(1976) is A_5. From Table 6, we can see that there is the fuzzy logical relationship $A_5(t-1) \rightarrow A_5(t), A_6(t)$ in Group 6. Based on Principle 2, the forecasted enrollment of year 1977 can be calculated as follows:

$$Forecasted = \frac{1*m_5+2*m_6}{1+2} = \frac{1*15401.75+2*15638.25}{3} = 1559.33$$

[1980] From Table 4, we can see that the fuzzified enrollments of years F(t-1)= F(1979) is A_9 . From Table 6, we can see that there is a fuzzy logical relationship $A_9(t-1) \rightarrow A_{11}(t)$, in Group 9. Based on Principle 1, the forecasted enrollment of year 1980 can be calculated as follows: $Forecasted = m_9 = 16715.5$

In the same way, we can get the forecasted enrollments of the other years of the University of Alabama based on the first-order fuzzy time series, as shown in Table 7.

Table 7: Forecasted enrollments of the proposed method using the first-order fuzzy time series.

Year	Actual	Fuzzified	Results
1971	13055	A1	Not forecasted
1972	13563	A1	13309
1973	13867	A2	13726.3
1974	14696	A3	14603.5
1975	15460	A5	15402
1976	15311	A5	15402
1977	15603	A6	15559.3
1978	15861	A7	15958.5
1979	16807	A9	16715.5
1980	16919	A11	17211.5
1981	16388	A8	16376.5
1982	15433	A5	15402
1983	15497	A5	15480.7
1984	15145	A4	15331.4
1985	15163	A4	15107.5
1986	15984	A7	15674.8
1987	16859	A10	16812.5
1988	18150	A12	18035
1989	18970	A13	18832
1990	19328	A14	19332

1991	19337	A14	19332
1992	18876	A13	18998.7

4. EXPERIMENTAL RESULTS.

The performance of the proposed method will be compared with the existing methods , such as the SCI model[2], the C96 model [4], the H01 model [5], CC06F model[11] and HPSO model [17] by using the enrollment of Alabama University from 1971 to 1992. It can be listed in Table 8 as below:

Table 8: A comparison of the forecasted results of the our proposed model with the existing models with first-order of the fuzzy time series under different number of intervals.

Year	Actual data	SCI	C96	H01	CC06F	HPSO	Our proposed
1971	13055						
1972	13563	14000	14000	14000	13714	13555	13309
1973	13867	14000	14000	14000	13714	13994	13726.3
1974	14696	14000	14000	14000	14880	14711	14603.5
1975	15460	15500	15500	15500	15467	15344	15402
1976	15311	16000	16000	15500	15172	15411	15402
1977	15603	16000	16000	16000	15467	15411	15559.3
1978	15861	16000	16000	16000	15861	15411	15958.5
1979	16807	16000	16000	16000	16831	16816	16715.5
1980	16919	16813	16833	17500	17106	17140	17211.5
1981	16388	16813	16833	16000	16380	16464	16376.5
1982	15433	16789	16833	16000	15464	15505	15402
1983	15497	16000	16000	16000	15172	15411	15480.7
1984	15145	16000	16000	15500	15172	15411	15331.4
1985	15163	16000	16000	16000	15467	15344	15107.5
1986	15984	16000	16000	16000	15467	16018	15674.8
1987	16859	16000	16000	16000	16831	16816	16812.5
1988	18150	16813	16833	17500	18055	18060	18035
1989	18970	19000	19000	19000	18998	19014	18832
1990	19328	19000	19000	19000	19300	19340	19332
1991	19337	19000	19000	19500	19149	19340	19332
1992	18876	19000	19000	19149	19014	19014	18998.7
MSE		423027	407507	226611	35324	22965	18770

Table 8 shows a comparison of the mean square error (MSE) of our method using the first-order fuzzy time series with different number of intervals, where the mean square error (MSE) is calculated as follows:

$$MSE = \frac{\sum_{i=1}^{N}(Fdi - Adi)^2}{N} = \frac{(13309-13563)^2 + (14603.5-14696)^2 ... + (13726.3-13867)^2}{21} \quad (9)$$

where N denotes the number of historical data in time series, Fd_i denotes the forecasted value at time i and Ad_i denotes the actual value at time i .

From Table 8, we can see that the proposed method has a smaller mean square error than than SCI model[2] the C96 model[4], the H01 model[5], the CC06F model[11] and HPSO model[17].

A comparison of the forecasted accuracy (i.e. the MSE value) between our model and CC06F model[11], HPSO model[17] under different number of intervals are listed in Table 9.

Table 9: A comparison of the forecasted accuracy between the our proposed method and HPSO model, the CC06F model under different number of intervals

Methods	Number of intervls						
	8	9	10	11	12	13	14
CC06F	132963	96244	85486	55742	54248	42497	35324
HPSO	119962	90527	60722	49257	34709	24687	22965
Our method	78950	42689	37265	35647	33834	21308	18770

From Table 9, it is clear that the our method is more powerful than the CC06 model and HPSO model with first-order of the fuzzy time series under different number of intervals

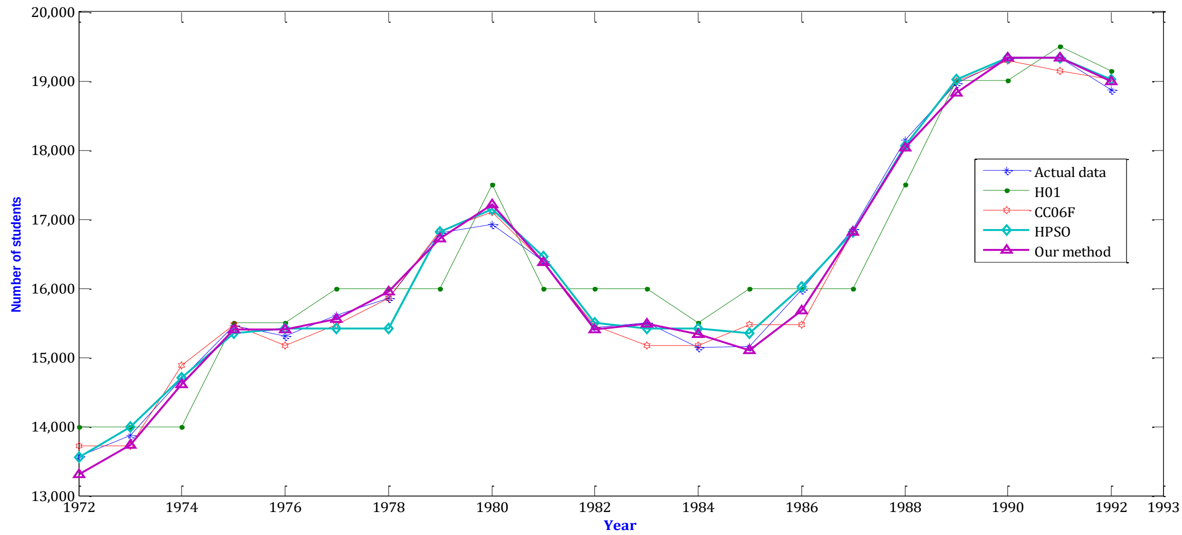

Fig-2: The curves of the actual data and the H01, CC06F, HPSO models and our model for forecasting enrollments of University of Alabama

Displays the forecasting results of CC06F model, HPSO model and the proposed method. The trend in forecasting of enrollment by first-order of the fuzzy time series in comparison to the actual enrollment can be visualized in Fig-2. In Fig-2, it can be seen that the forecasted value is close to the actual enrolment of students each year, from 1972s to 1992s

5. CONCLUSION

In this paper, we have proposed a new forecasting method in the first-order fuzzy time series model based on the time-variant fuzzy logical relationship groups and K-means clustering techniques. In this method, we tried to classify the historical data of Alabama University into clusters by K-means techniques and then, adjust the clusters into intervals with different lengths. In case study, we have applied the proposed method to forecast the number of students enrolling in the University of Alabama from 1972s to 1992s. The simulation result showed that the proposed method is able to obtain the forecasted value with better accuracy compared to other methods existing in literature. The detail of comparison was presented in Table 8 and 9.

Even the model was only examined in the enrolment forecasting problem; we believe that it can be applied to any other forecasting problems such as population, stock index, or market price forecasting, etc. That will be the future work of this research.

REFERENCES

[1] J.B. MacQueen, "Some methods for classi"cation and analysis of multivariate observations," in: Proceedings of the Fifth Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press, Berkeley, CA, pp. 281-297, 1967.

[2] Q. Song, B.S. Chissom, "Forecasting Enrollments with Fuzzy Time Series – Part I," Fuzzy set and system, vol. 54, pp. 1-9, 1993b.

[3] Q. Song, B.S. Chissom, "Forecasting Enrollments with Fuzzy Time Series – Part II," Fuzzy set and system, vol. 62, pp. 1-8, 1994.

[4] S.M. Chen, "Forecasting Enrollments based on Fuzzy Time Series," Fuzzy set and system, vol. 81, pp. 311-319. 1996.

[5] Hwang, J. R., Chen, S. M., & Lee, C. H. Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, 100(1–3), 217–228, 1998.

[6] Huarng, K. Heuristic models of fuzzy time series for forecasting. Fuzzy Sets and Systems, 123, 369–386, 2001b .

[7] Singh, S. R. A simple method of forecasting based on fuzzytime series. Applied Mathematics and Computation, 186, 330–339, 2007a.

[8] Singh, S. R. A robust method of forecasting based on fuzzy time series. Applied Mathematics and Computation, 188, 472–484, 2007b.

[9] S. M. Chen, "Forecasting enrollments based on high-order fuzzy time series", *Cybernetics and Systems: An International Journal,* vol. 33, pp. 1-16, 2002.

[10] H.K.. Yu "Weighted fuzzy time series models for TAIEX forecasting ", Physica A, 349 , pp. 609–624, 2005.

[11] Chen, S.-M., Chung, N.-Y. Forecasting enrollments of students by using fuzzy time series and genetic algorithms. International Journal of Information and Management Sciences 17, 1–17, 2006a.

[12] Chen, S.M., Chung, N.Y. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. International of Intelligent Systems 21, 485–501, 2006b.

[13] Liu, H.T., "An Improved fuzzy Time Series Forecasting Method using Trapezoidal Fuzzy Numbers," Fuzzy Optimization Decision Making, Vol. 6, pp. 63–80, 2007.

[14] Lee, L.-W., Wang, L.-H., & Chen, S.-M. Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms. Expert Systems with Applications, 33, 539–550, 2007.

[15] Jilani, T.A., Burney, S.M.A. A refined fuzzy time series model for stock market forecasting. Physica A 387, 2857–2862. 2008.

[16] Wang, N.-Y, & Chen, S.-M. Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series. Expert Systems with Applications, 36, 2143–2154, 2009.

[17] Kuo, I. H., Horng, S.-J., Kao, T.-W., Lin, T.-L., Lee, C.-L., & Pan. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. Expert Systems with applications, 36, 6108–6117, 2009a.

[18] S.-M. Chen, K. Tanuwijaya, " Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques", Expert Systems with Applications 38 ,15425–15437, 2011.

[19] Huarng, K.H., Yu, T.H.K., "*Ratio-Based Lengths of Intervals to Improve Fuzzy Time Series Forecasting*," IEEE Transactions on SMC – Part B: Cybernetics, Vol. 36, pp. 328–340, 2006.

BIOGRAPHIES

Nghiem Van Tinh is PhD student at Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam. He is working at Electronics Faculty, Thai Nguyen University of Technology, Vietnam. His research interests include Forecasting, clustering, Fuzzy logics, and Fuzzy time series.

Vu Viet Vu is a Doctor of Computer Science at Electronics Faculty, Thai Nguyen University of Technology. He has published some papers in machine learning topic in Pattern Recognition, and Intelligent Data Analysis journal. His research interests include machine learning, data mining, and pattern recognition

Tran Thi Ngoc Linh is a lecturer at Electronics Faculty, Thai Nguyen University of Technology, Vietnam. Her research interests include machine learning and data mining.