ENROLLMENTS FORECASTING BASED ON AGGREGATED K-MEANS CLUSTERING AND FUZZY TIME SERIES

Nghiem Van Tinh[1],Nguyen Thi Thu Hien[2] Nguyen Tien Duy[1]

[1]*Thai Nguyen University of Technology, Thai Nguyen University, Vietnam*
[2]*Thai Nguyen University of Education, Thai Nguyen University, Vietnam*

---***---

Abstract - *Most of the fuzzy forecasting methods based on fuzzy time series used the static length of intervals, i.e., the same length of intervals. The drawback of the static length of intervals is that the historical data are roughly put into intervals, even if the variance of the historical data is not high. In this paper, we present a new method for forecasting enrolments based on Fuzzy Time Series and K-Mean clustering(FTS-KM). To verify the effectiveness of the proposed model, the empirical data for the enrolments of the University of Alabama are illustrated, and the experimental results show that the proposed model outperforms those of previous some forecasting models with various orders and different interval lengths.*

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

1. INTRODUCTION

In the past decades, many methods have been presented for fuzzy forecasting [1]-[2], [6], [7]either to find a better forecasting result or to-do faster computations. The concept of fuzzy time series was proposed by Song and Chissom[1]-[3].However, most of the existing fuzzy forecasting methods based on fuzzy time series used the static length of intervals ,i.e., the same length of intervals such as [1-3], [4], [9],[10]. The drawback of the static length of intervals is that the historical data are roughly put into the intervals, even if the variance of the historical details not high. Moreover, the forecasting accuracy rates of the existing fuzzy forecasting methods based on the static length of intervals are not good enough. Recently, Ref.[16] presented a new hybrid forecasting model which combined particle swarm optimization with fuzzy time series to find proper length of each interval. Additionally, Ref.[17] proposed a new method to forecast enrolments based on automatic clustering techniques and fuzzy logical relationships. In this paper, a forecasting model based on two computational methods,K-mean clustering technique and fuzzy logical relationship groups 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.There are five sections in this paper. A brief introduction to fuzzy time series and K-mean clustering are given in Section 2. In Section 3, a FTS-KM based deterministic forecasting model is proposed and discussed. Then, the computational results are shown and analyzed in Section 4. The conclusions are discussed in Section 5.

2. FUZZY TIME SERIES AND K-MEANS CLUSTERING

2.1 Fuzzy time series

Fuzzy set theory was firstly developed by Zadeh in the 1965s to deal with uncertainty using linguistic terms. Ref.[1]successfully modelled the fuzzy forecast by adopting the fuzzy sets for fuzzy time series. To avoid complicated max–min composition operations, in[3]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→[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 $1\leq i\leq n$. General definitions offuzzy time series are given as follows:

Definition 1: **Fuzzy time series**

Let Y(t) (t =...,0, 1, 2 ...), a Sunset 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: 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 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 bybyF(t-λ), ..., F(t-2), F(t-1)→ F(t) and is called an λ- order fuzzy time series.

Definition 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 5: Fuzzy Relationship Group (FLRG)

Fuzzy logical relationships in the training datasets with the same fuzzy set on the left-hand-side can be further grouped into a fuzzy logical relationship groups. Suppose there are relationships such that

$A_i \rightarrow A_k$

$$A_i \rightarrow A_m$$

…….

So, based on [3], these fuzzy logical relationship can be grouped into the same FLRG as : $A_i \rightarrow A_k, A_m$ ...

2.2 K-means clustering technique

K-means clustering introduced by MacQueen [18] is one of the simplest unsupervised learning algorithms for solving the well-known clustering problem. The main idea of the K-means algorithm is the minimization of an objective function usually taken up as a function of the deviations between all patterns from their respective cluster centers[5].

3. A FORECASTING MODEL BASE ON K-MEANS CLUSTERING AND FUZZY TIME SERIES(FTS-KM)

An improved hybrid model for forecasting the enrolments of University of Alabama(named FTS-KM) based on the Fuzzy Time Series 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 create fuzzy relationship groups; and finally, we obtain the forecasting output based on the fuzzy relationship groups and rules of forecasting are our proposed. The historical data of enrolments of the University of Alabama are listed inTable 1.

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

Year	Actual	Year	Actual
1971	13055	1982	15433
1972	13563	1983	15497
1973	13867	1984	15145
1974	14696	1985	15163
1975	15460	1986	15984
1976	15311	1987	16859
1977	15603	1988	18150
1978	15861	1989	18970
1979	16807	1990	19328
1980	16919	1991	19337
1981	16388	1992	18876

Source:*In [1-3]*

3.1 Aclustering 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},{16 807},{16859},{16919},{18150},{18970,18876},{19328, 19337}

Step 2: Create the cluster enter and adjust the clusters into intervals.

In this step, we use automatic clustering techniques[17] to generate clustercenter($Center_j$) from clusters in step 1 according to qual(2).

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

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

Then, 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 $Ubound_j$of $cluster_j$and the lower bound $Lbound_{k+1}$of $cluster_{j+1}$can be calculated as follows:

$$Ubound_k = \frac{Center_k + Center_{k+1}}{2} \qquad (3)$$

$$Lbound_{k+1} = Cluster\_UB_k \qquad (4)$$

where k=1,...,p-1. Because there is no previouscluster before the first cluster and there is no next cluster after the last cluster, the lower bound$Lbound_1$of the first cluster and the upper bound $Ubound_p$of the last cluster can be calculated as follows:

$$Lbound_1 = \mathbf{Center_1} - (\mathbf{Center_1} - Cluster\_UB_1) (5)$$

$$Ubound_p = \mathbf{Center_p} + (\mathbf{Center_p} - Cluster\_LB_p) (6)$$

Table 2: Generate cluster center from clusters

No	Clusters	Center	Lbound	Ubound
1	{13055, 13563}	13309	13030	13588
2	{13867}	13867	13588	14434
3	14696, 15145, 15163,	15001	14434	15156
4	{15311}	15311	15156	15387
5	{15460, 15433, 15497}	15463	15387	15533
6	{15603}	15603	15533	15762.5
7	{15861, 15984}	15922	15762.5	16155
8	{16388}	16388	16155	16597.5
9	{16807}	16807	16597.5	16833
10	{16859}	16859	16833	16889
11	{16919}	16919	16889	17534.5
12	{18150}	18150	17534.5	18536.5
13	{18970, 18876}	18923	18536.5	19127.5
14	{19328, 19337}	19332	19127.5	19536.5

Step 3: Let each cluster $Cluster_j$form an interval $interval_j$, which means that the upper bound $Ubound_j$and the lower bound $Cluster_Lbound_j$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 $interval_j$according to [7] shown in Table 3:

$$Mid\_value_j = \frac{interval\_Lbound_j + interval\_Ubound_j}{2} (7)$$

where $interval_Ubound_j$and $interval_Lbound_j$are the upper bound and the lower bound of $interval_j$, respectively, and j= 1,…,p.

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

No	Intervals	M_j= Midpoint intevals
1	(13030, 13588]	13309
2	(13588, 14434]	14011
3	(14434, 15156]	14795
4	(15156, 15387]	15271.5
5	(15387, 15533]	15460
6	(15533, 15762]	15647.5
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 enrolments using FTS-KM

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

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

After applying the procedureK-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 enrolments 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 (8). 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 $a_{ij} \in [0,1]$, $1 \le i \le m$, and $1 \le j \le m$, $m=14$. Thevalue of a_{ij} indicates the grade of membership of u_j in the fuzzy set A_i. For example, the historical enrolment 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.

Table 4: Linguistic values for the enrolments 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	A4
1977	15603	A6
1978	15861	A7
1979	16807	A9
1980	16919	A11
1981	16388	A8
1982	15433	A5
1983	15497	A5
1984	15145	A3
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, 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.

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

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

Step 4: Establish and calculatethe forecasting values for all fuzzy logical relationship groups

By Chen [3], 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. Thus, from Table 5 and based on Definition 5, wecan obtain 14 fuzzy logicalrelationship groups and compute the forecasted output for these groups according to (9) and(10) are listed in Table 6.

Table 6:Fuzzy logical relationship groups(FLRGs)

Number of Groups	FLRGs	Forecasted value
1	A1 -> A1, A2	13660
2	A2 -> A3	14795
3	A3 -> A5, A4	15365.8
4	A5 -> A4, A5, A3	15175.5
5	A4 -> A6, A7	15803
6	A6 -> A7	15958.5
7	A7 -> A9, A10	16788.2
8	A9 -> A11	17211.5
9	A11 -> A8	16376.5
10	A8 -> A5	15460
11	A10 -> A12	18035
12	A12 -> A13	18832
13	A13 -> A14	19332
14	A14 -> A14, A13	19082

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

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

then the forecasted enrolment of year t forecasted =m_k (9)

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 enrolment of year t -1is $A_j$and there are the following fuzzy logical relationship group whose current state is A_j, shown as follows:

$A_j \rightarrow A_{i1}, A_{i2}, A_{ip}$

then the forecasted enrolment of year t is calculated as follows: $forecasted = \frac{\sum_{k=1}^{p} m_{ik}}{p}$ (10)

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

Step 5: Generate all fuzzy forecasting rules

Based on each group of fuzzy relationships created andrelative forecasting values in Step 4, we can generate correspondingfuzzy forecasting rules. The if-then statements are used as thebasic format for the fuzzy forecasting rules. Assume a first-orderfuzzy forecasting rule $R_i$is "if x = A, then y = B", the if-part of therule "x = A" is termed antecedent and the then-part of the rule"y = B" is termed consequent. For example, if we want to forecast enrolments Y(t) using fuzzy group 1 for the first-order time seriesin Table 6, the fuzzy forecasting rule$R_1$is will be "if $F(t-1) = A1$thenY(t) = 13660.

For example, Table 7demonstratesthe 14 fuzzy rules generated by the first-order fuzzy groups ofTable 6.

In the same way, we can get the 14 fuzzy rules based on the first-order fuzzy relationship groups, as shown in Table 7.

Table 7: The fuzzy if-then rules of the first-order fuzzy relationships on enrolments.

Rules	Antecedent	Consequent
1	If F(t-1)==A1	Then Y(t) =13660
2	If F(t-1)==A2	Then Y(t) =14795
3	If F(t-1)==A3	Then Y(t) =15365.8
4	If F(t-1)==A4	Then Y(t) =15175.5
5	If F(t-1)==A5	Then Y(t) =15803
6	If F(t-1)==A6	Then Y(t) =15958.5
7	If F(t-1)==A7	Then Y(t) =16788.2
8	If F(t-1)==A8	Then Y(t) =117211.5
9	If F(t-1)==A9	Then Y(t) =16376.5
10	If F(t-1)==A10	Then Y(t) =15460
11	If F(t-1)==A11	Then Y(t) =18023.5
12	If F(t-1)==A12	Then Y(t) =18778.5
13	If F(t-1)==A13	Then Y(t) =19212
14	If F(t-1)==A14	Then Y(t) =18995.2

Step 6: forecasting output based on the forecast rules

After the forecast rules are created, we can use them to forecast the training data. Suppose we want to forecast the data Y(t), we need to find out a matched forecast rule and get the forecasted value from this rule. If we use the first-order forecast rules listed in Table 6to forecast the data Y(t), we just simply find out the corresponding linguistic values ofF(t-1)with respect to the data Y(t-1) and then compare them to the matching parts of all forecast rules. Suppose a matching part of a forecast rule is matched, we then get a forecasted value from the forecasting part of this matched forecast rule. For example, if we want to forecast the data Y(1975), it is necessary to find out the corresponding linguistic values of F(1974) with respect to Y(1974). We then have the following pattern.

IfF(1974) == A3 then forecast Y(1975) = 15365.8 . In the same way, we complete forecasted results based on the first - order fuzzy forecast rules in Table 7are listed in Table 8.

Table 8:Forecasted enrolments of the first-order fuzzy relationships for Table 4.

Year	Actual	Fuzzified	Results
1971	13055	A1	Not forecasted
1972	13563	A1	13660
1973	13867	A2	13660
1974	14696	A3	14795
1975	15460	A5	15365.8
1976	15311	A4	15175.5
1977	15603	A6	15803
1978	15861	A7	15958.5
1979	16807	A9	16788.2
1980	16919	A11	17211.5
1981	16388	A8	16376.5
1982	15433	A5	15460
1983	15497	A5	15175.5
1984	15145	A3	15175.5
1985	15163	A4	15365.8
1986	15984	A7	15803
1987	16859	A10	16788.2

1988	18150	A12	18035
1989	18970	A13	18832
1990	19328	A14	19332
1991	19337	A14	19082
1992	18876	A13	19082

❖ Tocalculate the forecasted performance of proposed method in the fuzzy time series, the mean square error (MSE) is used as an evaluation criterion to represent the forecasted accuracy. The MSE value is computed as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(F_i - R_i)^2 \quad (11)$$

Where, $R_i$notes actual data on date i, F_i forecasted value on date i, n is number of the forecasted data

4. COMPUTATIONAL RESULTS.

The performance of the FTS-KM will be compared with the existing methods, such asthe **SCI** model[2],the **C96** model[3],the **H01** model[6] and **CC06F**model[11]by using the enrolment of Alabama University from 1971s to 1992s. It can be listed in Table 9as below:

Table 9:A comparison of the forecasted results of FTS-KMwith the existing models with first-order of the fuzzy time series under different number of intervals.

Year	Actual data	SCI	C96	H01	CC06F	FTS-KM
1971	13055	-	-	-	-	-
1972	13563	14000	14000	14000	13714	13660
1973	13867	14000	14000	14000	13714	13660
1974	14696	14000	14000	14000	14880	14795
1975	15460	15500	15500	15500	15467	15365.8
1976	15311	16000	16000	15500	15172	15175.5
1977	15603	16000	16000	16000	15467	15803
1978	15861	16000	16000	16000	15861	15958.5
1979	16807	16000	16000	16000	16831	16788.2
1980	16919	16813	16833	17500	17106	17211.5
1981	16388	16813	16833	16000	16380	16376.5
1982	15433	16789	16833	16000	15464	15460
1983	15497	16000	16000	16000	15172	15175.5
1984	15145	16000	16000	15500	15172	15175.5
1985	15163	16000	16000	16000	15467	15365.8
1986	15984	16000	16000	16000	15467	15803
1987	16859	16000	16000	16000	16831	16788.2
1988	18150	16813	16833	17500	18055	18035
1989	18970	19000	19000	19000	18998	18832
1990	19328	19000	19000	19000	19300	19332
1991	19337	19000	19000	19500	19149	19082
1992	18876	19000	19000	19149	19014	19082
MSE		**423027**	**407507**	**226611**	**35324**	**26113**

Table 9shows a comparison of MSE of our method using the first-order fuzzy time series under number of intervals=14, where MSE is calculatedaccording to (11) as follows:

$$MSE = \frac{\sum_{i=1}^{N}(F_i - R_i)^2}{N} = \frac{(13660-13563)^2 + (13660-13867)^2 \dots + (19082-18876)^2}{21} = 26113$$

where N denotes the number of forecasteddata, $F_i$denotes the forecasted value at time iand $R_i$denotes the actual value at time i.

From Table 9, we can see that the FTS-KM has a smaller mean square error than SCI model[2] the C96 model[3], the H01 model[6]and the CC06F model[11].

To verify the forecasting effectiveness for high-order FLRs, the C02model [9]isused to compare with the proposed model. From Table 10, The FTS-KM model gets the lowest MSE value of 20896 for the 8th-order FLRGs and The average MSE value is 35861.6 smaller than the C02 model.

Table 10:A comparison of the forecasted accuracy between our proposed method and C02 model, the CC06F model for seven intervals with different number of oders

Methods	Number of oders								
	2	3	4	5	6	7	8	9	Average
C02 model	89093	86694	89376	94539	98215	104056	102179	102789	**95868**
FTS-KM	82802	35767	27493	28141	29751	29269	20896	32231	**35861**

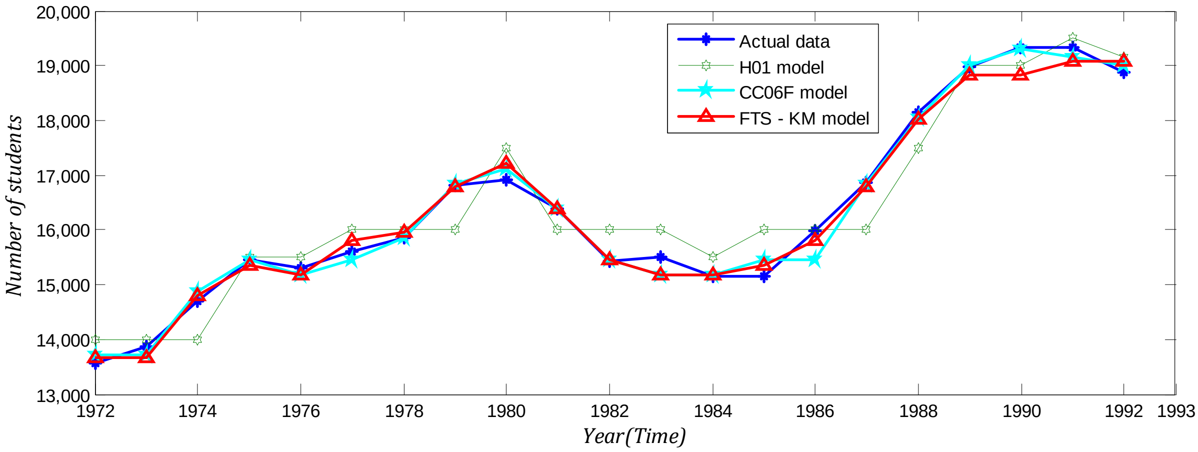

Fig-1: The curves of the actual data and the H01, CC06F, HPSO models and our model for forecasting enrollments of University of Alabama

Displays the forecasting results of H01 model, CC06Fmodel and our method. The trend in forecasting of enrolment by first-order of the fuzzy time series in comparison to the actual enrolment can be visualized in Fig-1.From Fig -1, the graphical comparison clearly shows that theforecasting accuracy of the proposed model is more precise than those of existing models with different first-order fuzzy logical relation.

5. CONCLUSION

In this paper, we have proposed a hybrid forecasting method in the first-order fuzzy time series model based on the time-invariant 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 9, Table 10 and Fig-1.

Although this study shows the superior forecasting capability compared with existing forecasting models; but the proposed model is a new forecasting model and only tested by the enrolment data. To assess the effectiveness of the forecasting model, there are two suggestions for future research: The first, we can apply proposed model to deal with more complicated real-world problems for decision-making such as weather forecast, crop production, stock markets, and etc. The second, we use more intelligent methods (e.g., particle swarm optimization, ant colony or a neural network) to deal with forecasting problems. That will be the future work of this research.

REFERENCES

[1] Q. Song, B.S. Chissom, "Forecasting Enrollments with Fuzzy Time Series – Part I," Fuzzy set and system, vol. 54, pp. 1-9, 1993b.
[2] Q. Song, B.S. Chissom, "Forecasting Enrollments with Fuzzy Time Series – Part II," Fuzzy set and system, vol. 62, pp. 1-8, 1994.
[3] S.M. Chen, "Forecasting Enrollments based on Fuzzy Time Series," Fuzzy set and system, vol. 81, pp. 311-319. 1996.
[4] 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.
[5] Michael K. Ng, "A note on constrained k-means algorithms," Pattern Recognition, vol. 13, pp. 515-519, 2000.
[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 enrolments using high-order fuzzy time series and genetic algorithms. International of Intelligent Systems 21, 485–501, 2006b.
[13] 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.
[14] Jilani, T.A., Burney, S.M.A. A refined fuzzy time series model for stock market forecasting. Physica A 387, 2857–2862. 2008.
[15] 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.
[16] 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.
[17] 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.
[18] J.B. MacQueen, "Some methods for classication 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.

BIOGRAPHIES

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

Nguyen Tien Duy is PhD student at Institute of Information Technology, Vietnam Academy of Science and Technology, Hanoi, Vietnam.He is working at Electronics Faculty, Thainguyen University of Technology, Thainguyen, Vietnam. Research interests: Engineering controls, Fuzzy logics, Fuzzy control and Computing.

Nguyen Thi Thu Hien is a lecturer at Faculty of Mathematics, Thai Nguyen University of Education, Vietnam. Her research interests include data mining and machine learning.