A FORECASTING METHOD BASED ON COMBINING AUTOMATIC CLUSTERING TECHNIQUE AND FUZZY RELATIONSHIP GROUPS

Nghiem Van Tinh

Thai Nguyen University of Technology, Thai Nguyen University
Thai Nguyen, Vietnam

---

*Abstract – Over the past few years, some methods have been presented based on fuzzy time series(**FTS**) to forecast real problems, such as forecasting stock market, forecasting enrolments, temperature prediction, etc. When forecasting these problems based on fuzzy time series, it is obvious that the length of intervals in the universe of discourse is important because it can affect the forecasting accuracy rate. However, some of the existing fuzzy forecasting methods based on fuzzy time series used the static length of intervals, i.e., the same length of intervals. The disadvantage of the static length of intervals is that the historical data are put into the intervals in a rough way, even if the change of the historical data is not large. Therefore, the forecasting accuracy rates of the existing fuzzy forecasting methods are not good enough. Consequently, we need to propose a new fuzzy forecasting method to overcome the drawbacks of the existing forecasting models to increase the forecasting accuracy rates. In this paper, a hybrid forecasting model based on two computational methods, the fuzzy logical relationship groups and clustering technique, is presented for forecasting enrolments. Firstly, we use the automatic clustering algorithm to divide the historical data into clusters and adjust them into intervals with unequal 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 and high – order fuzzy time series, our method gets a higher average forecasting accuracy rate than the existing methods.*

Key Words: *Fuzzy time series, forecasting, Time-invariant fuzzy relationship groups, clustering, enrolments.*

1. INTRODUCTION

It can be seen that forecasting activities play an important role in our daily life. Therefore, many more forecasting models have been developed to deal with various problems in order to help people to make decisions, such as crop forecast [6], [7] academic enrolments [1], [9], the temperature prediction [13], stock markets [14], 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. [1], [2]] proposed the time-invariant FTS and the time-variant FTS model

which use the max–min operations to forecast the enrolments of the University of Alabama. However, the main drawback of these methods is enormous computation load. Then, Ref. [3] proposed the first-order FTS model by introducing a more efficient arithmetic method. After that, FTS has been widely studied to improve the accuracy of forecasting in many applications. Ref. [4] considered the trend of the enrolment in the past years and presented another forecasting model based on the first-order FTS. Ref. [12] pointed out that the effective length of the intervals in the universe of discourse can affect the forecasting accuracy rate. In other words, the choice of the length of intervals can improve the forecasting results. Ref. [5] presented a heuristic model for fuzzy forecasting by integrating Chen's fuzzy forecasting method [3]. At the same time, Ref.[8] proposed several forecast models based on the high-order fuzzy time series to deal with the enrolments forecasting problem. In [9], the length of intervals for the FTS model was adjusted to forecast the Taiwan Stock Exchange (TAIEX)

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 two new methods to forecast enrolments, temperature and TAIFEX forecasting based on automatic clustering techniques and high – order fuzzy logical relationships.

In this paper, we proposed a hybrid forecasting model combining the time-invariant fuzzy relationship groups and automatic clustering technique in [18] .

In case study, we applied the proposed method to forecast the enrolments of the University of Alabama. Computational results show that the proposed model outperforms other existing methods.

Rest of this paper is organized as follows. The fundamental definitions of FTS and automatic clustering technique are discussed in Section 2. In Section 3, we use an automatic clustering algorithm combining the FTS for forecasting the enrolments of the University of Alabama. In Section 4 presents the results from the application of the proposed method to real data sets. Then, the computational results are shown and analyzed in Section 5. Finally, conclusions are presented in Section 6.

2. FUZZY TIME SERIES AND AUTOMATIC CLUSTERING ALGORITHM

In this section, we provide briefly some definitions of fuzzy time series in Subsection 2.1 and Automatic clustering algorithm in Subsection 2.2

2.1. FTS definitions

General definitions of fuzzy time series are given as follows:

Definition 1: Fuzzy time series [[1], [2], [3]]

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, . . .).

[Example] The common observations of daily weather condition for certain region can be described using the daily common words “hot”, “very hot”, “cold”, “very cold”, “good”, “very good”, etc. All these words can be represented by fuzzy sets.

Definition 3: Fuzzy logic relationship [[1], [2], [3]]

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 presented by fuzzy logical relationship A_i → 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 4: λ- order fuzzy time series [8]

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 $\boldsymbol{\lambda}$- order fuzzy time series.

Definition 5: Fuzzy Relationship Group (FLRG) [3]

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

$$A_i \rightarrow A_{j1}$$
$$A_i \rightarrow A_{j2}$$
....

So, based on Chen[3] these fuzzy logical relationships can be grouped into the same FLRG as : $A_i \rightarrow A_{j1}, A_{j2} ...$

2.2. An automatic clustering algorithm

A cluster is a set whose elements have the similar properties in some sense. Elements in the same cluster have the same properties while elements in different clusters have different properties. If the elements in a cluster are numerical values, then the smaller the distance (i.e., the difference) between two elements in the cluster, the higher the degree of similarity between these two elements.

In this section, we briefly summarize an automatic clustering algorithm to divived the numerical data into clusters. The algorithm is introduced in [18]. The algorithm is composed of the main following steps.

1. Sort the numerical data in an ascending order.
   $d_1, d_2, d_3, \ldots, d_i, \ldots, d_n$. with $d_{i-1} < d_i$
   where d_1 is the smallest datum among the n numerical data, d_n is the largest datum among the n numerical data, and $1 \leq i \leq n$
2. Calculate the average distance *aver_dif* of the distances between every pair of neighboring numerical data in the sorted data sequence

Based on the value of *aver_dif,* determine wherever two adjacent numerical data d_i and d_j in the data sequence can be put into the current cluster or needs to be put it into a new cluster

3. FORECASTING MODEL BASED ON COMBINED AUTOMATIC CLUSTERING AND FUZZY TIME SERIES

An improved hybrid model for forecasting the enrolments of University of Alabama based on Automatic clustering technique and FTS. At first, we apply automatic 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 relationships 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 step-wise procedure of the proposed model is detailed as follows: Step 1: Creating intervals from historical data of enrolments based on automatic clustering algorithm

Assume that the clusters are obtained in Subsection 2.2 as following:

$$\{d_1, d_2\}; \{d_3, d_4\}; \{d_5, d_6\}; \ldots; \{d_k\}; \ldots; \{d_{n-1}, d_n\}.$$

Transform these clusters into contiguous intervals based on the following Sub-steps:

Step 1.1*: Transform the first cluster* $\{d_1, d_2\}$ *into the interval* $[d_1, d_2)$

Step 1.2*: set* $\{d_1, d_2\}$ *is the current interval and let* $\{d_3, d_4\}$ *is the current cluster*

begin

if $(d_2 \geq d_3)$ **then**

begin

transform *the current cluster* $\{d_3, d_4\}$ *into interval* $[d_2, d_4)$

set $[d_2, d_4)$ *as the current interval, and set the next cluster* $\{d_5, d_6\}$ *as the current cluster*

end;

if $(d_2 < d_3)$ **then**

begin

transform $\{d_3, d_4\}$ *into interval* $[d_3, d_4)$ *create a new interval* $[d_2, d_3\}$ *between* $[d_1, d_2)$ *and* $[d_3, d_4)$

set $[d_3, d_4)$ *is the current interval, and set the next cluster* $\{d_5, d_6\}$ *as the current cluster.*

end;

.....

if *the current interval is* $[d_i, d_j)$ *and the current cluster is* $\{d_k\}$ **then**

begin

transform *the current interval* $[d_i, d_j)$ *into interval* $[d_i, d_k)$ *set* $[d_i, d_k)$ *as the current interval, and set the next cluster as the current cluster.*

end; **end**.

Step 1.3*: Repeatedly check the current interval and the current cluster until all the clusters have been transformed into intervals*

Step 2: Define the fuzzy sets for each interval

Assume that there are n intervals $u_1, u_1, u_1, ..., u_n$ for data set obtained in Step 1. For n intervals, there are n linguistic values which are $A_1, A_2, A_3, \dots, A_{n-1}$ and A_n to represent different regions in the universe of discourse, respectively. Each linguistic variable represents a fuzzy set A_i $(1 \le i \le n)$ and its definition is described in (1) .

$$A_i = \sum_{j=1}^{7} \frac{a_{ik}}{u_k} \quad (1)$$

where $a_{ik} \in [0,1]$, $1 \le i \le n$, $1 \le k \le n$ and u_k is the k-th interval. The value of a_{ik} indicates the grade of membership of u_j in the fuzzy set A_i and it is shown as following:

$$a_{ik} = \begin{cases} 1 & \text{if } k == i \\ 0.5 & \text{if } k == i-1 \text{ or } k == i+1 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

Step 3: Fuzzify variations of the historical enrolment data

In order to fuzzify all historical data, it's necessary to assign a corresponding linguistic value to each interval first. The simplest way is to assign the linguistic value with respect to the corresponding fuzzy set that each interval belongs to with the highest membership degree.

Step 4: Identify all fuzzy relationships

Relationships are identified from the fuzzified historical data obtained in Step 3. If the fuzzified enrollments of years t and t - 1 are A_i and A_j, respectively, then construct the first – order fuzzy logical relationship "$A_i \rightarrow A_j$", where A_i and A_j are called the fuzzy set on the left-hand side and fuzzy set on the right-hand side of fuzzy logical relationships, respectively.

Step 5: Construct the 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. Suppose there are relationships such that

$A_i \rightarrow A_{j1}$; $A_i \rightarrow A_{j2}$; ....

We can be grouped into a relationship group as follows: $A_i \rightarrow A_{j1}, A_{j2} \dots$

Step 6: Calculate the forecasted outputs.

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

Rule 1: If the fuzzified enrolment 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: $A_j \rightarrow A_k$;then the forecasted enrolment of year t forecasted = 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

Rule 2: If the fuzzified enrolment 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:

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

then the forecasted enrolment of year t is calculated as follows: $forecasted = \frac{m_1 + m_2 + \dots + m_p}{p}$; wiht $p \le n$

where $m_1, m_2, \dots and\ m_p$ are the middle values of the intervals u_1 , u_2 and u_p respectively, and the maximum membership values of $A_1, A_2, \dots, A_p$ occur at intervals $u_1, u_2, \dots, u_p$, respectively.

Rule 3: If the fuzzified enrolment of year t 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: $A_j \rightarrow \#$

where the symbol "#" denotes an unknown value, then the forecasted enrollment of year t + 1 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 u_j.

4. THE APPLICATION TO ENROLLMENT DATA

To verify the effectiveness of the proposed model, all historical enrolments in Table 1 (*the enrolment data at the University of Alabama from 1971s to 1992s*) are used to illustrate for forecasting process. The step-wise procedure of the proposed model is presented as following:

Table 1: Historical enrolments of the University of Alabama

Year	Actual data	Year	Actual data
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

Step 1: After applying the automatic clustering algorithm for clustering the historical numerical data, we can get 21 intervals which are shown in Table 2.

Table 2. Intervals obtained from automatic clustering

No	Intervals	No	Intervals
1	u_1 = [13055, 13354.1]	12	u_{12} =[15984, 16088.9]
2	u_2 = [13354.1, 13862.1]	13	u_{13} = [16088.9, 16687.1]
3	u_3 = [13862.1, 14166.1]	14	u_{14} = [16687.1, 16807]
4	u_4 = [14166.1, 14396.9]	15	u_{15} = [16807, 16919]
5	u_5 = [14396.9, 14995.1]	16	u_{16} = [16919, 17850.9]
6	u_6 = [14995.1, 15145]	17	u_{17} = [17850.9, 18449.1]
7	u_7 = [15145, 15163]	18	u_{18} = [18449.1, 18876]
8	u_8 = [15163, 15311]	19	u_{19} = [18876, 18970]
9	u_9 = [15311, 15603]	20	u_{20} = [18970, 19328]
10	u_{10} = [15603, 15861]	21	u_{21} = [19328, 19337]
11	u_{11} = [15861, 15984]		

Step 2: Define fuzzy sets for each interval

For 21 intervals, there are 21 linguistic values which are $A_1, A_2, A_3, \dots, A_{n-1}$ and A_n, shown as follows:

$$\begin{aligned} A_1 &= \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \dots + \frac{0}{u_{21}} \\ A_2 &= \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \dots + \frac{0}{u_{21}} \\ &\text{--------------------------------} \\ A_{21} &= \frac{0}{u_1} + \frac{0}{u_2} + \dots + \frac{0.5}{u_{20}} + \frac{1}{u_{21}} \end{aligned} \quad (3)$$

If the historical data belongs to ui, where $(1 \le i \le 21)$, then the datum is fuzzified into A_i. For example, from Table 1, we can see that the historical data of year 1971 is 13055, where 13055 falls in the interval u = [13055, 13354.1). Therefore, the enrolment of year 1971 (i.e., 13055) is fuzzified into A_1. The results of fuzzification are listed in Table 3, where all historical data are fuzzified to be fuzzy sets.

Table 3: Fuzzified enrolments of the University of Alabama

Year	Actual data	Fuzzy set	Year	Actual data	Fuzzy set
1971	13055	A1	1982	15433	A9
1972	13563	A2	1983	15497	A9
1973	13867	A3	1984	15145	A7
1974	14696	A5	1985	15163	A8
1975	15460	A9	1986	15984	A12
1976	15311	A9	1987	16859	A15
1977	15603	A10	1988	18150	A17
1978	15861	A11	1989	18970	A20
1979	16807	A15	1990	19328	A21
1980	16919	A16	1991	19337	A21
1981	16388	A13	1992	18876	A19

Step 3: Identify all fuzzy relationships

From Table 3 and base on Definition 3, we get first – order fuzzy logical relationships are shown in Table 4.

Table 4: The first- order fuzzy logical relationship

No	Fuzzy relations	No	Fuzzy relations
1	A1 -> A2	11	A13 -> A9
2	A2 -> A3	12	A9 -> A7
3	A3 -> A5	13	A7 -> A8
4	A5 -> A9	14	A8 -> A12
5	A9 -> A9	15	A12 -> A15
6	A9 -> A10	16	A15 -> A17
7	A10 -> A11	17	A17 -> A20
8	A11 -> A15	18	A20 -> A21
9	A15 -> A16	19	A21 -> A21
10	A16 -> A13	20	A21 -> A19

Step 4: Establish all fuzzy logical relationship groups

From Table 4 and based on Definition 5, we can obtain 14 fuzzy relationship groups, as shown in Table 5.

Table 5: The first- order fuzzy relationship groups

No	Relationships	No	Relationships
1	A1 -> A2	9	A16 -> A13
2	A2 -> A3	10	A13 -> A9
3	A3 -> A5	11	A7 -> A8
4	A5 -> A9	12	A8 -> A12
5	A9 -> A9, A10, A7	13	A12 -> A15
6	A10 -> A11	14	A17 -> A20
7	A11 -> A15	15	A20 -> A21
8	A15 -> A16, A17	16	A21 -> A21, A19

Step 5: Calculate the forecasting value by using the three rules following as.

For example, the forecasted enrollments of the years 1972 and 1980 are calculated as follows

[**1972**] From Table 3, we can see that the fuzzified enrollment of year 1972 is A_2. From Table 4, we can see that there is a fuzzy logical relationship "A_1 -> A_2" in Group 1. Therefore, the forecasted enrollment of year 1972 is equal to the middle value of the interval u_2. Because the middle value of the interval u_2 is 13608, the forecasted enrollment of 1972 is 13608.

[**1980**] From Table 3, we can see that the fuzzified enrollment of year 1980 is A_{16}. From Table 4, we can see that there is a fuzzy logical relationship "A_{15} -> A_{16}, A_{17}" in Group 8. Therefore, the forecasted enrollment of year 1980 is calculated as follows:

Forecasted $= \frac{m_{16}+m_{17}}{2} = \frac{17384.95+18150}{2} = 17767.48$

where 17384.95 and 18150 are the middle values of the intervals u_{16} and u_{17}, respectively.

In the same way, the other forecasted enrollments of the University of Alabama based on the first-order fuzzy time series are listed in Table 6.

Table 6: Forecasted enrolments of the proposed method using the first-order fuzzy time series.

Year	Actual	Fuzzified	Results
1971	13055	A1	Not forecasted
1972	13563	A2	13608.1
1973	13867	A3	14014.1
1974	14696	A5	14696
1975	15460	A9	15457
1976	15311	A9	15447.7
1977	15603	A10	15447.7
1978	15861	A11	15922.5
1979	16807	A15	16863
1980	16919	A16	17767.5
1981	16388	A13	16388
1982	15433	A9	15457
1983	15497	A9	15447.7
1984	15145	A7	15447.7
1985	15163	A8	15237
1986	15984	A12	16036.4
1987	16859	A15	16863
1988	18150	A17	17767.5
1989	18970	A20	19149
1990	19328	A21	19332.5
1991	19337	A21	19127.8
1992	18876	A19	19127.8

To evaluate the forecasted performance of proposed method in the FTS, the mean square error (MSE) and the mean absolute percentage error (MAPE) are used as a comparison criterion to represent the forecasted accuracy. The MSE value and MAPE value are computed according to (4) and (5) as follows:

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

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{F_i - R_i}{R_i}\right| * 100\%$$

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

5. EXPERIMENTAL RESULTS

Experimental results for our model will be compared with the existing methods, such as the **SCI** model [2] the **C96** model [3] and the **H01** model [5] by using the enrolment of Alabama University from 1972s to 1992s are listed in Table 7 .

Table 7 shows a comparison of MSE and MAPE of our method using the first-order FTS under different number of intervals, where MSE and MAPE are calculated according to (4) and (5) as follows:

$$MSE = \frac{\sum_{i=1}^{21}(F_i - R_i)^2}{N} = \frac{(13608-13563)^2+(14014.1-13867)^2 ... +(19127-18876)^2}{21} = 56297$$

$$MAPE = \frac{1}{21}\sum_{i=1}^{21}\left|\frac{F_i - R_i}{R_i}\right| * 100\% = \frac{1}{21}(\frac{abs(13608-13563)}{13563} + \cdots + \frac{abs(19127-18876)}{18876}) = 0.849\%$$

where N denotes the number of forecasted data, F_i denotes the forecasted value at time i and R_i denotes the actual value at time i.

Table 7: A comparison of the forecasted results of the 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	Our model
1971	13055	-	-	-	-
1972	13563	14000	14000	14000	13608.1
1973	13867	14000	14000	14000	14014.1
1974	14696	14000	14000	14000	14696
1975	15460	15500	15500	15500	15457
1976	15311	16000	16000	15500	15447.7
1977	15603	16000	16000	16000	15447.7
1978	15861	16000	16000	16000	15922.5
1979	16807	16000	16000	16000	16863
1980	16919	16813	16833	17500	17767.5
1981	16388	16813	16833	16000	16388
1982	15433	16789	16833	16000	15457
1983	15497	16000	16000	16000	15447.7
1984	15145	16000	16000	15500	15447.7
1985	15163	16000	16000	16000	15237
1986	15984	16000	16000	16000	16036.4
1987	16859	16000	16000	16000	16863
1988	18150	16813	16833	17500	17767.5
1989	18970	19000	19000	19000	19149
1990	19328	19000	19000	19000	19332.5
1991	19337	19000	19000	19500	19127.8
1992	18876	19000	19000	19149	19127.8
MSE		**423027**	**407507**	**226611**	**56297**
MAPE		**3.22%**	**3.11%**	**2.66%**	**0.85%**

From Table 7, we can see that the proposed method has a smaller MSE value of 56297 and MAPE value of 0.85% than SCI model the C96 model [3] and the H01 model [5]. To verify the forecasting effectiveness for high-order FLRs methods in the C02 model [8], CC06b model [11] shown in Table 8, are selected to compare with our model. From Table 8, the proposed model also gets the lowest MSE value of **16143** for the 3rd-order FLRGs among all the compared models and The average MSE value is **18419.5** smaller than the C02 model and CC06b model.

To be clearly visualized, Fig. 1 depicts the trends for actual data and forecasted results of the H01 model with forecasted results of proposed method. From Fig. 1, It is obvious that the forecasting accuracy of the proposed model is more close than any existing models for the first-order fuzzy logical relationships with different number of intervals.

In addition, Displays the forecasting results of C02 model, CC06b model and our model. The trend in forecasting of enrolment by high-order of the fuzzy time series in comparison to the actual enrolment can be visualized in Fig.2.

Table 8: A comparison of the forecasted accuracy between the proposed method and, the C02 model, CC06b model under different number of intervals with various number of orders by the MSE value.

Methods	Number of oders							
	2	3	4	5	6	7	8	9	Average(*MSE*)
C02 model	89093	86694	89376	94539	98215	104056	102179	102789	**95868**
CC06b model	67834	31123	32009	24948	26980	26969	22387	18734	**31373**
Our model	21300	**16143**	17040	18042	17837	17917	18931	20146	**18419.5**

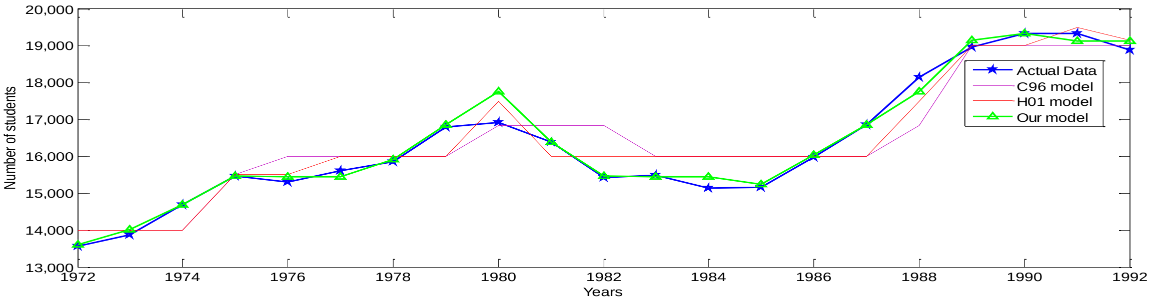

Fig. 1: The curves of the H01 models and our model for forecasting enrolments of University of Alabama

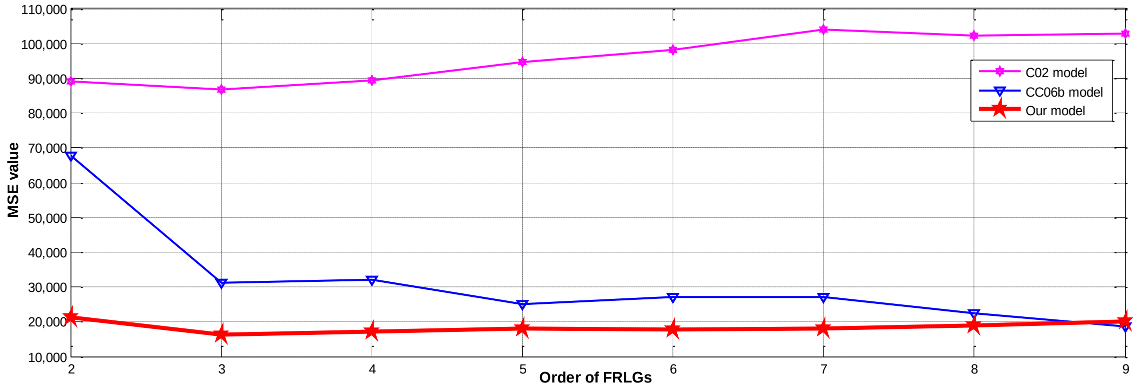

Fig. 2: A comparison of the MSE values for different intervals with the high-order FTS model.

6. CONCLUSION

In this paper, we have presented a hybrid forecasted method to handle forecasting enrolments of the University of Alabama based on the fuzzy logical relationship groups and automatic clustering techniques. Firstly, the proposed method applies the automatic clustering algorithm to divide the historical data into clusters and adjust them into intervals with different lengths. Secondly, we fuzzify all the historical data of the enrolments and establish the fuzzy relation groups. Thirdly, we calculate forecasting output and compare forecasting accuracy with other existing models. Lastly, based on the performance comparison in Tables 7, 8 and Fig. 1, 2, it can show that our model outperforms previous forecasting models for the training phases with various orders and different interval lengths.

The proposed model was only tested by the forecasting enrollment problem, and it can actually be applied to other practical problems such as population forecast, and stock index forecast in the further research.

ACKNOWLEDGMENT

The work reported in this paper has been supported by the Science Council of Thai Nguyen University of Technology - Thai Nguyen University, under contract number 110xHĐNCKH-ĐHKTCN

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

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

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

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

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

[10]. 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.

[11]. 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.

[12]. Huarng, K. theH. Effective lengths of intervals to improve forecasting fuzzy time series. Fuzzy Sets and Systems, 123(3), 387–394, 2001a.

[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]. Shyi-Ming Chen, Nai-Yi Wang Jeng-Shyang Pan. Forecasting enrollments using automatic clustering techniques and fuzzy logical relationships, Expert Systems with Applications 36 , 11070–11076, 2009.

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.