A Proposed Model For Forecasting Stock Markets Based On Clustering Algorithm And Fuzzy Time Series

Nghiem Van Tinh
Thai Nguyen University of Technology, Thai Nguyen University
Thai Nguyen, Vietnam

***Abstract*—The forecast of fuzzy time series has been widely applied to many various fields such as enrollments, stocks market, weather, population growth prediction and so on. In recent years, many researchers used fuzzy time series to handle prediction problems. When forecasting these problems, 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 (the same length of intervals). The drawback 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 huge. Therefore, the forecasting accuracy rates of the existing fuzzy forecasting methods are not good enough. Thus, 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 proposed clustering algorithm, is presented for forecasting the Taiwan Futures Exchange (TAIFEX). Firstly, we use the proposed 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 TAIFEX and calculate the forecasted output by the proposed model.**

The experimental results and the comparison results show that the proposed method can be successfully applied in stock market forecasting or similar kinds of time series. Additionally, compared to the other existing methods based on the first-order and high – order fuzzy time series, our method outperformed these compared methods

Keywords—Fuzzy time series(FTS), forecasting, fuzzy logical relationships (FLRs), clustering, TAIFEX.

I. INTRODUCTION

Song and Chissom [1], [2] proposed the time invariant fuzzy time series model and the time-variant fuzzy time series model to deal with the forecasting problems in which the historical data are represented by linguistic values. Because Song and Chissom's used the max–min operations to forecast the enrollments of the University of Alabama, they take a lot of computation time to deal with max– min composition operations. Chen [3] presented a method to forecast the enrollments of the University of Alabama

by using a simple fuzzy time series forecasting model. In recent years, some researchers focused on the research topic of using fuzzy time series for handling forecasting problems, such as forecasting the enrollments [4] - [8], crop forecast [9], [10], stock index prediction [11], [12], the temperature prediction [12]. Chen also extended his previous work [3] to present several forecast models based on the high-order FTS to deal with the enrollments forecasting problem [5], [13]. Huarng [7] pointed out that the different lengths of intervals in the universe of discourse can affect the forecasting result and a proper choice of the length of each interval can greatly improve the forecasting accuracy rate. In other words, the choice of the length of intervals can improve the forecasting results. Ref. [8] presented a heuristic model for fuzzy forecasting by integrating Chen's fuzzy forecasting method [3]. In [14] the length of intervals for the FTS model was adjusted by the K-mean clustering algorithm to forecast the enrollments. Some other techniques for determining proper intervals and interval lengths is used automatic clustering technique [15], [16] and particle swarm optimization algorithm [17], [18]. Additionally, in [19] proposed a new method to forecast Temperature and TAIFEX based on automatic clustering algorithm and two – factors high – order fuzzy time series.

In this paper, a hybrid forecasting model based on combining the clustering algorithm for partitioning the universe of discourse and the fuzzy logical relationship groups is presented. Firstly, we apply proposed clustering technique to classify the collected data into clusters and adjust these clusters into contiguous intervals. secondly, based on the new intervals obtained, the proposed method fuzzifies the historical data into fuzzy sets to form fuzzy logical relationships. Then, it calculates the value of the variable between the subscripts of adjacent fuzzy sets appearing in the antecedents of fuzzy logical relationships. Then, it lets the fuzzy logical relationships with the same variable value form a high-order fuzzy logical relationship group. Finally, it chooses a fuzzy logical relationship group for forecasting.

In case study, we applied the proposed method to forecast the Taiwan Futures Exchange. Computational results show that the proposed model outperforms other existing methods.

This paper is organized as following. The fundamental definitions of FTS and proposed clustering technique are discussed in Sect 2. In Sect 3, we use the clustering technique combining the FTS for forecasting the TAIFEX. The computational results are shown and analyzed in Sect 4. Finally, conclusions are presented in Sect 5

II. FUZZY TIME SERIES AND CLUSTERING ALGORITHM

In this section, we provide briefly some definitions of fuzzy time series [1], [2] in Subsect A and proposed clustering algorithm in Subsect B.

A. Fuzzy Time Series

In [2] , Song and Chissom proposed the definition of fuzzy time series based on fuzzy sets ,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, $f_A : U \to [0,1]$, $f_A(u_i)$ 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 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 3: 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 $A_i \to$ 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

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 Groups (FLRGs)

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

$A_i \to A_j$

$A_i \to A_k$

…….

So, these fuzzy logical relationships can be grouped into the same FLRG as : $A_i \to A_j, A_k \ldots$

B. The proposed clustering algorithm

In this section, we briefly summarize an automatic clustering algorithm to divide the numerical data into clusters. A simple data clustering algorithm used in this research is as follows:

1. Sort the numerical data in an ascending order. Assume that the ascending sequence of the data is shown as follows: $\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \ldots, \mathrm{d}_\mathrm{i}, \ldots, \mathrm{d}_\mathrm{n}$. 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 $1 \le i \le n$.

2. Based on the ascending data sequence calculate the average distance difference $Aver_{diff}$ between any two adjacent data and calculate the standard deviation $\boldsymbol{Dev_{diff}}$ of the difference between any two adjacent data, shown as follows:

$$Aver_{diff} = \frac{\sum_i^{n-1}(d_{i+1}-d_i)}{n-1} \quad (1)$$

$$Dev_{diff} = \sqrt{\frac{\sum_i^{n-1}(d_{i+1}-d_i-d)^2}{n-1}} \quad (2)$$

where d denotes the mean of the data $\mathrm{d}_1, \mathrm{d}_2, \mathrm{d}_3, \ldots, \mathrm{d}_\mathrm{i}, \ldots,$ and d_n

$$d = \frac{\sum_{i=1}^{n} d_i}{n} \quad (3)$$

3. Create a new cluster for the first datum, i.e. the smallest datum, and let the cluster to be the current cluster. Calculate the maximum distance *Max_Twodata* between any two adjacent data using $\boldsymbol{Dev_{diff}}$, shown as follows:

$$Max\_Twodata = 0.5 * \boldsymbol{Dev_{diff}} \quad (3)$$

Based on the value of *Max_Twodata,* determine whether the following datum can be put into the current cluster or needs to be put it into a new cluster. Assume that there is a d_i cluster as follows:

$\ldots, \{\ldots, \mathrm{d}_\mathrm{i}\}, \mathrm{d}_{\mathrm{i}+1}, \mathrm{d}_{\mathrm{i}+2}, \ldots, \mathrm{d}_\mathrm{n}$, where $(1 \le i \le n)$

if ($\mathrm{d}_{i+1} - \mathrm{d}_\mathrm{i}$) < Max_Twodata**) then**

put d_i into the current cluster in which di belongs

else

Create a new cluster for d_{i+1} and let the new cluster in which d_{i+1} belongs be the current cluster.

If ($\mathrm{d}_{i+1} - \mathrm{d}_i$ < *max_data_distance)*, is true for all the data set), then find the mean value of the maximum and minimum difference value and use it to calculate the maximum data distance to create the clusters.

$$Max\_Twodata = 0.5 * \boldsymbol{Aver_{diff}} \quad (4)$$

Otherwise subtract the minimum value difference min(**(** $\mathrm{d}_{i+1} - \mathrm{d}_\mathrm{i}$)) between any data in the data set from the maximum value different max **((** $\mathrm{d}_{i+1} - \mathrm{d}_\mathrm{i}$)) between any data in the data set and divide by two, use the result to calculate maximum data distance

$$Max\_Twodata = 0.5 * (max(\mathrm{d}_{i+1} - \mathrm{d}_\mathrm{i}) - \min(\mathrm{d}_{i+1} - \mathrm{d}_\mathrm{i})) \quad (5)$$

Repeatedly perform the above process, until all the data have been clustered.

4. Adjust the clusters into intervals according to the follow rules:

Assume that there are two adjacent clusters, $cluster_i$ and $cluster_k$ shown as follows:

$\ldots, \{d_{i1}, \ldots, d_{in}\}, \{d_{k1}, \ldots, d_{km}\}, \ldots$; where d_{in} is the last datum in $cluster_i$ and d_{k1} is the first datum in $cluster_k$

Then the upper bound *Cluster _ uBᵢ* of *clusterᵢ* and the lower bound *cluster_lBₖ* of *clusterₖ* can be calculated as follows:

$$Cluster\_uB_i = \frac{d_{in} + d_{k1}}{2} \quad (6)$$

$$Cluster\_lB_k = Cluster\_uB_i \quad (7)$$

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 $Cluster\_uB_n$ of the last cluster can be calculated as follows:

$$Cluster\_lB_1 = d_1 - \text{Max}\_Twodata \quad (8)$$

$$Cluster\_uB_n = d_n + \text{Max}\_Twodata \quad (9)$$

where d_1 is the first datum in the first cluster and d_n is the last datum in the last cluster.

The clusters themselves correspond to intervals, where the *upper bound and the lower bound of an interval* are taken from *the upper bound and the lower bound of a cluster*, respectively.

Calculate the middle value Mid_value_k of the interval $interval_k$ as follows:

$$Mid\_value_i = \frac{interval\_lB_i + interval\_uB_i}{2} \quad (10)$$

where $interval\_lB_i$ and $interval\_UB_i$ denote the lower bound and the upper bound of the interval $interval_i$, respectively, with i = 1,..,n.

III. Forecasting Model Based On Clustering Algorithm And Fuzzy Time Series

An improved hybrid model for forecasting the Taiwan Futures Exchange (TAIFEX) based on clustering technique and FTS. Firstly, we apply 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 logical relationships and create fuzzy logical relationship groups and finally, we obtain the forecasting output based on the fuzzy relationship groups and rules of forecasting are our proposed.

To verify the effectiveness of the proposed model, all historical TAIFEX [8] (*Historical data of the TAIFEX under 8/3/1998 – 8/31/1998)* in Table 1 are used to illustrate for forecasting process. The step-wise procedure of the proposed model is detailed as follows:

Table 1: Historical data of the TAIFEX [8]

Date	Actual	Date	actual
8/3/1998	7552	8/18/1998	7220
8/4/1998	7560	8/19/1998	7285
8/5/1998	7487	8/20/1998	7274
8/6/1998	7462	8/21/1998	7225
8/7/1998	7515	8/24/1998	6965
8/10/1998	7365	8/25/1998	6949
8/11/1998	7360	8/26/1998	6790
8/12/1998	7330	8/27/1998	6835
8/13/1998	7291	8/28/1998	6695
8/14/1998	7320	8/29/1998	6728
8/15/1998	7300	8/31/1998	6566
8/17/1998	7219		

Step 1: *Creating intervals from historical data of TAIFEX based on proposed clustering algorithm.*

In this step, the clustering algorithm in subsection A of Part II is applied for creating intervals. After applying the procedure proposed clustering, we can get the following 21 intervals and the value of the intervals are shown in Table 2.

Table 2: The partition from Taifex data

No	Intervals	No	Intervals
1	(6566, 6611.2)	10	(7225, 7274)
2	(6611.2, 6695)	11	(7274, 7300)
3	(6695, 6728)	12	(7300, 7320)
4	(6728, 6790)	13	(7320, 7330)
5	(6790, 6835)	14	(7330, 7360)
6	(6835, 6949)	15	(7360, 7365)
7	(6949, 6965)	16	(7365, 7462)
8	(6965, 7219)	17	(7462, 7487)
9	(7219, 7225)	18	(7487, 7560)

Step 2: Define the fuzzy sets for each interval

Assume that there are n intervals $u_1, u_1, u_1, \ldots, u_n$ for data set obtained in Step 1. For n intervals, there are n linguistic values which are $A_1, A_2, A_3, \ldots, A_{n-1}$ 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 (11) .

$$A_i = \frac{a_{i1}}{u_1} + \frac{a_{i2}}{u_2} + \cdots + \frac{a_{ij}}{u_j} + \cdots + \frac{a_{in}}{u_n} \quad (11)$$

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

$$a_{ij} = \begin{cases} 1 & \text{if } j == i \\ 0.5 & \text{if } j == i-1 \text{ or } j == i+1 \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

For 19 intervals, there are 19 linguistic values which are $A_1, A_2, A_3, \ldots, A_{35}$ and A_{36}, shown as follows:

$$\begin{aligned} A_1 &= \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \cdots + \frac{0}{u_{19}} \\ A_2 &= \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \cdots + \frac{0}{u_{19}} \\ &\text{-----------------------------------} \\ A_{36} &= \frac{0}{u_1} + \frac{0}{u_2} + \cdots + \frac{0.5}{u_{18}} + \frac{1}{u_{19}} \end{aligned} \quad (13)$$

Step 3: Fuzzy all historical TAIFEX dataset

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. For example, the historical data on date 8/3/1998 is 7552, and it belongs to interval u_{18} because 7552 is within (7487, 7560]. So, we then assign the linguistic value A_{18} corresponding to interval u_{18} to it. Consider two time serials data $Y(t)$ and $F(t)$ on date t, where $Y(t)$ is actual data and $F(t)$ is the fuzzy set of $Y(t)$. According to Eq. (13), the fuzzy set A_{18} has the maximum membership value at the interval u_{18}. Therefore, the historical data time series on date Y(8/3/1998) is fuzzified to A_{18}. The completed fuzzified results of the TAIFEX are listed in Table 3.

TABLE 3: THE RESULTS OF FUZZIFICATION

Date	Actual data	linguistic value
8/3/1998	7552	A18
8/4/1998	7560	A18
8/5/1998	7487	A17
8/6/1998	7462	A17
8/7/1998	7515	A18
8/10/1998	7365	A16
8/11/1998	7360	A15
8/12/1998	7330	A14
8/13/1998	7291	A11
8/14/1998	7320	A13
8/15/1998	7300	A12
8/17/1998	7219	A9
8/18/1998	7220	A9
8/19/1998	7285	A11
8/20/1998	7274	A11
8/21/1998	7225	A10
8/24/1998	6965	A8
8/25/1998	6949	A7
8/26/1998	6790	A5
8/27/1998	6835	A6
8/28/1998	6695	A3
8/29/1998	6728	A4
8/31/1998	6566	A1

Step 4: Identify all λ - order fuzzy relationships

Based on Definition 3 and 4. To establish a λ - order fuzzy relationship, we should find out any relationship which has the $F(t-\lambda), F(t-\lambda+1), \ldots, F(t-1) \rightarrow F(t)$, where $F(t-\lambda), F(t-\lambda+1), \ldots, F(t-1)$ and $F(t)$ are called the current state and the next state, respectively. Then a λ - order fuzzy relationship in the training phase is got by replacing the corresponding linguistic values. For example, supposed $\lambda = 1$ from Table 3, a fuzzy relation $A_{18} \rightarrow A_{18}$ is got as $F(8-3-1998) \rightarrow F(8-4-1998)$. So on, we get 21 the first-order fuzzy relationships are shown in Table 4.

TABLE 4: THE FIRST-ORDER FUZZY LOGICAL RELATIONSHIPS

No	Fuzzy relations	No	Fuzzy relations
1	A18 -> A18	12	A9 -> A11
2	A18 -> A17	13	A11 -> A11
3	A17 -> A18	14	A11 -> A10
4	A18 -> A16	15	A10 -> A8
5	A16 -> A15	16	A8 -> A7
6	A15 -> A14	17	A7 -> A5
7	A14 -> A11	18	A5 -> A6
8	A11 -> A13	19	A6 -> A3
9	A13 -> A12	20	A3 -> A4
10	A12 -> A9	21	A4 -> A1
11	A9 -> A9		

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 first – order fuzzy logical relationships such that

$A_i \rightarrow A_j$; $A_i \rightarrow A_k$; ....

We can be grouped into a relationship group as follows: $A_i \rightarrow A_j\, A_k \ldots$

Where, the fuzzy logical relationships as the same counted only once. By the same way for establishing high – order fuzzy logical relationships.

From Table 4 and based on Definition 5, we can obtain 16 first – order fuzzy relationship groups, as shown in Table 5.

TABLE 5: THE FIRST-ORDER FUZZY LOGICAL RELATIONSHIP GROUPS

No	Relationships	No	Relationships
1	A18 -> A18, A17, A16	9	A9 -> A9, A11
2	A17 -> A18	10	A10 -> A8
3	A16 -> A15	11	A8 -> A7
4	A15 -> A14	12	A7 -> A5
5	A14 -> A11	13	A5 -> A6
6	A11 -> A13, A11, A10	14	A6 -> A3
7	A13 -> A12	15	A3 -> A4
8	A12 -> A9	16	A4 -> A1

Step 6: Calculate defuzzified the forecasted output

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

Rule 1: If the fuzzified TAIFEX on date 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 TAIFEX of date 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 TAIFEX of date 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 TAIFEX of date t is calculated as follows: $forecasted = \frac{m_1+m_2+\cdots+m_p}{p}$; $p \le n$

where $m_1, m_2, \ldots and\ m_p$ 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 , ... ,A_p occur at intervals u_1 , u_2,…, u_p , respectively.

Rule 3: If the fuzzified TAIFEX of date 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 TAIFEX of date 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.

For example, the forecasted TAIFEX on the dates of month 8/4/1998 and 8/12/1998 are calculated as follows:

[**8/3/1998**] From Table 3, we can see that the fuzzified TAIFEX of F(8/4/1998) is A_{18}. From Table 5, we can see that there is a fuzzy logical relationship "A_{18} -> A_{18}" in Group 1 as follows: A18 -> A18, A17, A16. Therefore, the forecasted on date 8/3/1998 is calculated as follows:

$Forecasted = \frac{m_{18}+m_{17}+m_{16}}{3} = \frac{7523.5+7474.5+7413.5}{3} =$ 7470.5; where 7523.5, 7474.5 and 7413.5 are the middle values of the intervals u_{18} , u_{17} and u_{16} , respectively.

[**8/12/1998**] From Table 3, we can see that the fuzzified TAIFEX of F(8/12/1998) is A_{14}. From Table 5, we can see that there is a fuzzy logical relationship

"A_{15} -> A_{14}" in Group 4. Therefore, the forecasted TAIFEX of 8/12/1998 is calculated as follows:
Forecasted(8/12/1998) = $m_{14} = 7345$
where 7345 is the middle values of the intervals u_{14}.
In the same way, the other forecasted TAIFEX based on the first-order fuzzy time series are listed in Table 6:

TABLE 6: THE COMPLETE FORECASTED TAIFEX FROM 8/3/1998 TO 8/31/1998 BASED ON FIRST-ORDER FUZZY TIME SERIES

Date	Actual data	linguistic value	Forecasted value
8/3/1998	7552	A18	
8/4/1998	7560	A18	7470.5
8/5/1998	7487	A17	7470.5
8/6/1998	7462	A17	7470.5
8/7/1998	7515	A18	7523.6
8/10/1998	7365	A16	7470.5
8/11/1998	7360	A15	7362.5
8/12/1998	7330	A14	7345
8/13/1998	7291	A11	7287
8/14/1998	7320	A13	7287.2
8/15/1998	7300	A12	7310
8/17/1998	7219	A9	7222
8/18/1998	7220	A9	7254.5
8/19/1998	7285	A11	7254.5
8/20/1998	7274	A11	7287.2
8/21/1998	7225	A10	7287.2
8/24/1998	6965	A8	7092
8/25/1998	6949	A7	6957
8/26/1998	6790	A5	6812.5
8/27/1998	6835	A6	6892
8/28/1998	6695	A3	6711.5
8/29/1998	6728	A4	6759
8/31/1998	6566	A1	6588.6

To evaluate the forecasted performance of proposed method in the FTS, the mean square error (MSE) is used as a comparison criterion to represent the forecasted accuracy. The MSE value are calculated according to (14) as follows:

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

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

IV. EXPERIMENTAL RESULTS

In this paper, we apply the proposed method to forecast TAIFEX index with the whole historical data [8], from 8/3/1998 to 9/30/1998 are used to perform comparative study in the training phase. In order to verify the forecasting effectiveness of the proposed model with the high-order FTS and different numbers of intervals . Experimental results for our model will be compared with the existing methods, such as the **C96** [3] model, **H01a** [7] model, **H01b** [8] model, **L07** [12] model and **L06** 0 model based on the first – order and high-order FTS are listed in Table 7.

Table 7 shows a comparison of MSE of our method using the high-order fuzzy logical relationship groups under different number of intervals, where MSE value are calculated according to (14) as follows:

$$MSE = \frac{\sum_{i=1}^{N}(F_i - R_i)^2}{N} = \frac{(7345-7330)^2+(7287-7291)^2 \ldots +(6796.5-6787)^2}{40} = 186.18$$

TABLE 7: A COMPARISON OF FORECASTING VALUES OF THE PROPOSED MODEL WITH EXISTING MODEL BASED ON FIRST ORDER AND 7TH –ORDER FTS

Date	Actual data	C96[3]	H01a[7]	H01b[8]	L06[20]	L07[12]	Our model	
							7^{th} - order	*First- order*
08/03/1998	7552	-----	----	-----	-----	-----	-----	
08/04/1998	7560	7450	7450	7450	-----	-----	-----	7535.9
08/05/1998	7487	7450	7450	7450	-----	-----	-----	7535.9
08/06/1998	7462	7500	7450	7500	7450	-----	-----	7444
08/07/1998	7515	7500	7500	7500	7550	-----	-----	7515.8
08/10/1998	7365	7450	7450	7450	7350	-----	-----	7444
08/11/1998	7360	7300	7350	7300	7350	-----	-----	7362.5
08/12/1998	7330	7300	7300	7300	7350	7348	7345	7345
08/13/1998	7291	7300	7350	7300	7250	7301.5	7287	7287
08/14/1998	7320	7183.33	7100	7188.33	7350	7311.5	7325	7287.2
08/15/1998	7300	7300	7350	7300	7350	7301.5	7310	7310
08/17/1998	7219	7300	7300	7300	7250	7226.5	7222	7222
08/18/1998	7220	7183.33	7100	7100	7250	7226.5	7222	7254.5
08/19/1998	7285	7183.33	7300	7300	7250	7301.5	7287	7254.5
08/20/1998	7274	7183.33	7100	7188.33	7250	7256.5	7287	7287.2
08/21/1998	7225	7183.33	7100	7100	7250	7226.5	7249.5	7287.2
08/24/1998	6955	7183.33	7100	7100	6950	6952	6987.2	6987.2
08/25/1998	6949	6850	6850	6850	6950	6952	6930.8	6930.8
08/26/1998	6790	6850	6850	6850	6750	6783.5	6796.5	6859
08/27/1998	6835	6775	6650	6775	6850	6852	6838.5	6838.5
08/28/1998	6695	6850	6750	6750	6650	6713	6702.4	6761.4
08/29/1998	6728	6750	6750	6750	6750	6713	6745	6735.1
08/31/1998	6566	6775	6650	6650	6550	6561	6566	6566
09/01/1998	6409	6450	6450	6450	6450	6406	6416.6	6416.6
09/02/1998	6430	6450	6550	6550	6450	6406	6483.2	6592.8
09/03/1998	6200	6450	6350	6350	6250	6198.5	6215.3	6215.3
09/04/1998	6403.2	6450	6450	6450	6450	6406	6416.6	6416.6
09/05/1998	6697.5	6450	6550	6550	6650	6703	6702.4	6592.8
09/07/1998	6722.3	6750	6750	6750	6750	6713	6725.2	6735.1
09/08/1998	6859.4	6775	6850	6850	6850	6852	6861.5	6821.1
09/09/1998	6769.6	6850	6750	6750	6750	6783.5	6768.3	6865.7
09/10/1998	6709.75	6775	6650	6650	6750	6713	6716	6823.4
09/11/1998	6726.5	6775	6850	6775	6750	6713	6725.2	6725.2
09/14/1998	6774.55	6775	6850	6775	6817	6783.5	6780.8	6821.1

Journal of Multidisciplinary Engineering Science Studies (JMESS)
ISSN: 2458-925X
Vol. 3 Issue 9, September - 2017

09/15/1998	6762	6775	6650	6775	6817	6783.5	6768.3	6768.3
09/16/1998	6952.75	6775	6850	6850	6817	6953	6930.8	6823.4
09/17/1998	6906	6850	6950	6850	6950	6952	6930.8	6859
09/18/1998	6842	6850	6850	6850	6850	6852	6847	6859
09/19/1998	7039	6850	6950	6850	7050	7089	7039	7039
09/21/1998	6861	6850	6850	6850	6850	6852	6861.5	6861.5
09/22/1998	6926	6850	6950	6850	6950	6952	6930.8	6865.7
09/23/1998	6852	6850	6850	6850	6850	6852	6861.5	6859
09/24/1998	6890	6850	6950	6850	6850	6893	6898	6865.7
09/25/1998	6871	6850	6850	6850	6850	6852	6880.5	6880.5
09/28/1998	6840	6850	6750	6750	6850	6852	6838.5	6838.5
09/29/1998	6806	6850	6750	6850	6850	6792.5	6820.5	6761.4
09/30/1998	6787	6850	6750	6750	6750	6783.5	6796.5	6796.5
MSE		***9668.94***	***7856.5***	***5437.58***	***1364.56***	***249.61***	***186.18***	***2538***

From Table 7, we can see that the proposed method has the MSE value of 2538 smaller than the methods presented in C96 [3], H01b [7] and H01b [8] model for the first-order fuzzy time series with different number of intervals. In addition, the proposed method obtains a lowest average MSE value of 186.18 among two forecasting models presented in L06 0 and L07 [12] for the high-order fuzzy logical relationship groups with different number of intervals.

V. Conclusions

In this paper, we have presented a hybrid forecasted method to handle forecasting the TAIFEX based on the fuzzy logical relationship groups and proposed clustering algorithm. Firstly, using proposed clustering algorithm to divide the historical data into clusters and adjust them into intervals with different lengths. Our model found out proper lengths of intervals in the universe of discourse. Secondly, we fuzzify all the historical data of the enrolments and establish the fuzzy logical relationship groups. Thirdly, we calculate forecasting output and compare forecasting accuracy with other existing models. Lastly, From the experimental study on the TAIFEX forecasting, the results have shown that the proposed model has higher forecasting accuracy than some compared models with various orders and different interval lengths.

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 TAIFEX dataset. 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, earthquake forecasting and etc. The second, we can use type-2 fuzzy time series or hegde algebras to deal with more complex forecasting problems. That will be the future work of this research.

References

[1] Song, Q., & Chissom, B. S. Fuzzy time series and its model. Fuzzy Sets and Systems, 54(3), 269–277, 1993a.

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

[6] 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.

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

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

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

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

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

[12] 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

[13] 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.

[14] N. V. Tinh, et al. "A Forecasting Model Based on K-Means Clustering and Time-Invariant Fuzzy Relationship Groups", Journal of Multidisciplinary Engineering Science Studies, Vol 2, Issue 8, 815 -821, 2016

[15] 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.

[16] 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.

[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] Huang, Y. L. et al. A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization. Expert Systems with Applications, 38, 8014–8023, 2011.

[19] 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

Lee, L. W., Wang, L. H., Chen, S. M., & Leu, Y. H. Handling forecasting problems based on two-factors high-order fuzzy time series. IEEE Transactions on Fuzzy Systems, 14(3), 468–477, 2006.