International Journal of Multidisciplinary and Current Research ISSN: 2321-3124
Research Article Available at: http://ijmcr.com

An Interval Partitioning Approach Using Graph–Based Clustering in Fuzzy Time Series Forecasting Model

Nghiem Van Tinh[1,*], Tran Thi Thanh[1], Bui Thi Thi[1]

[1]Thai Nguyen University of Technology, Thai Nguyen University, ThaiNguyen, VietNam

Received 01 Feb 2021, Accepted 01 April 2021, Available online 06 April 2021, **Vol.9 (March/April 2021 issue)**

Abstract

In the implementations of fuzzy time series (FTS) forecasting model, the determination of interval lengths has an important impact on the performance of the forecasting model. However, Equal length intervals used in most existing literatures are convenient but have been chosen arbitrarily. Huarng developed a new approach which is called distribution- and average-based length in order to determine effective length of partitioned intervals. In this study, a new FTS forecasting model which uses a graph- based clustering technique to determine different length of intervals is proposed. The proposed forecasting model has been applied to the two time series data, which are well-known enrolments data at The University of Alabama and numerical data sets of Gasonline prices in VietNam. The computational results show that the proposed model gets a higher forecasting accuracy than the existing models when is applied to enrolments data at The University of Alabama.

Keyword. *Enrolments, forecasting, FTS, fuzzy relationships, clustering technique*

Introduction

Forecasting the future of any event or phenomenon plays a vital role in our day-to-day life. It helps to make better decisions under uncertain situations. However, forecast of future values of these events with full accuracy is very difficult, but their forecasting accuracy and the speed of forecasting process can be enhanced. Previously, the most popular linear forecasting models that have been successfully applied in practical applications are the regression analysis model, exponential moving average model and autoregressive integrated moving averages (ARIMA) model. Unfortunately, the disadvantages of the traditional time series forecasting methods are that (1) they cannot deal with forecasting problems in which historical data are linguistic values or uncertainty and (2) their forecasting accuracy rates are not good enough they face. To solve this problem, Song and Chissom [1] proposed a model in 1993 based on ambiguity and estimated knowledge incorporated in the data of time series. Firstly, they gave a model by the help of fuzzy sets [2] which represent or take care of all these ambiguities, and called this concept as "Fuzzy Time Series (FTS)". In 1996, Chen [3] developed FTS model based on first-order fuzzy logical relationships (FLRs), and obtained the forecasted results with simplified arithmetic operations rather than complicated max-min composition operations. Chen's [3] forecasting results were far better than the models proposed by Song and Chissom [1, 4, 5]. Recently, many studies provided some improvements in Chen's [3] model in terms of finding effective lengths of intervals [6–10], fuzzification of time series data set [11], establishment of FLRs [12], and defuzzification [13] To improve the forecasting accuracy, just now various FTS models are proposed by many researchers. For example, Chen et al. [14] introduced a new FTS model for stock price forecasting by using the theory of the Fibonacci sequence. This model is based on the framework of the conventional FTS models, whose forecasting accuracy is outperformed than these models [4, 15]. Recently, researchers in these articles [16–20] introduced computational methods for forecasting by which high order FLR's are blown away the fault of first-order FTS models [3, 21]. To minimize the time of complicated computations of fuzzy relational matrices or to find the steady state of fuzzy relational matrix, Singh [22] proposed a new method in FTS modeling approach. Li and Cheng [23] introduced a new fuzzy deterministic model to solve three major issues, viz., to restraint the ambiguity in forecasting, to divide the intervals adequately, and to achieve the forecasted accuracy with various interval lengths. For these purposes, they designed an optimized FTS forecasting procedure by which the forecasted data is treated as a trapezoidal fuzzy number rather than a single point data. The proposed model gives better forecasting results than the conventional FTS models [3, 11, 24]. In [25], Singh and Borah proposed two-

*Corresponding author's ORCID ID: 0000-0000-0000-0000
DOI: https://doi.org/10.14741/ijmcr/v.9.2.6

factors high-order FTS model to resolve the problem associated with determination of lengths of intervals and data defuzzification. The proposed model is based on the hybridization of artificial neural network (ANN) with FTS. The comparative analyzes signify that this model exhibits higher accuracy than those of existing two-factors models [26–31]. In [32], Singh and Borah presented a new model based on hybridization of FTS theory with ANN. In this model, an ANN based architecture is incorporated to defuzzify the fuzzified time series values and to obtain the forecasting results. Chen [33] presented a novel high-order FTS model to solve nonlinear problem of time series data set. In this model, entropy-based clustering technique and ANN based architecture are employed to overcome the problems of data partition and representation of FLRs, respectively.

As above-mentioned, determining the suitable length of intervals, establishing fuzzy relationships considered to be challenging tasks and significant influencing the accuracy of forecasting model. In this research, we present a new forecasting model which uses a graph- based clustering technique to determine different length of intervals on enrollment data at The University of Alabama. In this approach, initially, we proposed a new algorithm for finding the best interval lengths based on a graph - based clustering algorithm. Then, we define various fuzzy sets based on these evolved intervals and fuzzy the historical data into fuzzy sets. Based on these fuzzified values, we derive the FLRs. Then, we obtain weighted fuzzy relationship groups according to their chronological order (FLRGs) from the FLRs. Later, all these FLRGs are used to obtain the forecasting results based on the weighted defuzzification method.

The rest of this paper is organized as follows: Basic definitions of fuzzy time series and algorithms are given in Sec2. Section 3 presents a forecasting model which combines with the FTS and Graph-based clustering algorithm. Section 4 evaluates the models' performance and compares obtained results to those of other models. Finally, Section 5 provides some conclusions

2. Basic Concepts of FTS and Algorithms

2.1. Basic concepts of FTS

The definition of FTS was originally introduced by Song and Chissom [1, 4]. Some concepts associated with fuzzy time series are recalled in here. Let $U = \{u_1, u_2, \dots, u_n\}$ be an universe of discourse (UoD); a fuzzy set A of U can be defined as $A = \{\mu_A(u_1)/u_1 +, \mu_A(u_2)/u_2 \dots + \mu_A(u_n)/u_n\}$, where μ_A :U→[0,1] is the membership function of A, $\mu_A(u_i)$ indicates the degree of membership of u_i in the fuzzy set A, $f_A(u_i) \in [0, 1]$, and $1 \le i \le n$.

Definition 1: Fuzzy time series [1]

Let $Y(t)$ $(t = .., 0, 1, 2 ..)$, a subset of R, be the UoD on which fuzzy sets $f_i(t)$ $(i = 1,2 \dots)$ are defined and if $F(t)$ is a collection of $f_1(t), f_2(t), \dots$, then $F(t)$ is called a FTS definition on $Y(t)$ (t . . ., 0, 1,2 . ..).

Definition 2: Fuzzy logical relationships (FLRs**)** [1, 3]

Support that $F(t) = A_j$ and $F(t-1) = A_i$, the first – order relationship between F(t-1) and F(t) is referred as a FLR: $A_i \rightarrow A_j$. Where, A_j is caused by A_i and A_i, A_j is called the left - hand side and the right-hand side of FLR, respectively.

Definition 3: m - order fuzzy logical relationships [16]

If $F(t)$ is caused by previous states $F(t-1)$, $F(t-2), \dots, F(t-m+1)$ $F(t-m)$ then this fuzzy logical relationship is represented by $F(t-m), \dots, F(t-2)$, $F(t-1) \rightarrow F(t)$ and it is called an m - order FLR. Time series with respect to this FLR is called an m – order FTS.

Definition 4: Fuzzy relationship groups (FRGs) [3]

The fuzzy logical relationships having the same left- hand side can be further grouped into a FRG. Assume there are exists FLRs as follows: $A_i \rightarrow A_{k1}$, $A_i \rightarrow A_{k2}$,.., $A_i \rightarrow A_{km}$; these FLRs can be put into the same FRG as : $A_i \rightarrow A_{k1}, A_{k2}, \dots, A_{km}$.

Definition 5: Time-variant fuzzy relationship groups [34]

The fuzzy logical relationship is determined by the relationship F(t-1)→ F(t). If, let $F(t) = A_i(t)$ and $F(t-1) = A_j(t-1)$, The FLR between F(t-1) and F(t) can be denoted as $A_j(t-1) \rightarrow A_i(t)$. Also at the time t, we have the following fuzzy logical relationships: $A_j(t-1) \rightarrow A_i(t)$, $A_j(t1-1) \rightarrow A_{i1}(t1), .., A_j(tn-1) \rightarrow A_{in}(tn)$ with $t1, t2, .., tn \le t$. It is noted that $A_i(t_1)$ and $A_i(t_2)$ with the same A_i but appear at different times t_1 and t_2, respectively. It means that if these FLRs occur before $A_j(t-1) \rightarrow A_i(t)$, we can group the FLRs having the same left-hand side into a FRG as $A_j(t-1) \rightarrow A_{i1}(t1), A_{i2}(t2), A_{in}(tn), A_i(t)$. It is called first – order TV-FRGs.

2.2. Graph – based clustering algorithm

In this study, we introduce a data clustering method which is a class of graph-based method to represent the time series data set into clusters. The proposed clustering method displays the dataset in the form of a tree and automatically generates clusters instead of the number of clusters pre-selected by the user. The proposed clustering method displays the dataset in the form of a tree and automatically generates clusters instead of the number of clusters pre-selected by the user. In particular, in this paper, the graph-based clustering method is introduced by an algorithm including four procedures as follows:

(1) The Procedure of Finding Root Node (PFRN). Based on the input data, this procedure points out the root node.

(2) Node Insertion Procedure (NIP). This procedure inputs one element of data set and root node, and set the elements in the tree at their proper position.
(3) Tree Creation Procedure (TCP). From the input data set and the root node, this procedure shows the tree.
(4) Node Clustering Procedure (NCP). This procedure inputs the tree which is generated by the TCP, and makes logical clustering of the nodes.

The Graph-based clustering algorithm

```
Input: S(x_1, x_2, ...x_n)
Output: Clusters C(c_1, c_2, ...c_k )
BEGIN
(1) PROCEDURE_PFRN (S)
BEGIN
// Calculate range (Rg) based on Maximum and Minimum value of S
Rg = MAX_value – MIN_value
// Calculate standard deviation (SD) of the S.
For each i=1 to N
{
   Mean = average(X_i)
   SD = Sqrt(1/i sum(X_i − Mean)^2)
}
  w = R_g / (SD*N)
// Define UoD (U) of the S
U = [MIN_value – w, MAX_value + w];
//Calculate midpoint of U as:
Mid_u = (MIN_value+ MAX_value) / 2
Root = Mid_u
END;
------------------------------------------------------
(2) PROCEDURE_NIP (Root, S)
BEGIN
if (X_i < Root) then
if (Root.LEFT < > NULL) then
   Call:  NIP(Root. LEFT, X_i)
else
Root.LEFT = NULL
end if
else if (X_i> Root) then
    if (Root. RIGHT < > NULL) then
   Call: NIP(Root. RIGHT, X_i)
    else
   Root. RIGHT = NULL
end if
END;
------------------------------------------------------
(3) PROCEDURE_TCP (Root, S)
BEGIN
For each i = 1 to N
NIP(Root, X_i)
END;
------------------------------------------------------
(4) PROCEDURE_NCP (Root)
BEGIN
if (Root == NULL) then
{
   "No Tree Found"
return
}
else if (Root.RIGHT < > NULL && Root.LEFT < > NULL) then
if (Root is not presented in Cluster) then
{
  minDiffnode = makeDiff(Root, Root. RIGHT, Root. LEFT);
// the difference between values of the Root and Root.RIGHT, and Root and Root.LEFT
makeCluster(Root, minDiffnode)
}
if (minDiffnode == Root.RIGHT) then
if ((Root.RIGHT).LEFT < > NULL) then

add (Root.RIGHT).LEFT ; // add child node in the Cluster
end if
if ((Root.RIGHT). RIGHT < >NULL) then
Call: NCP((Root.RIGHT).RIGHT)
end if
Call: NCP(Root.LEFT)
else
if ((Root.LEFT). LEFT < >NULL) then
Call: NCP((Root.LEFT).LEFT)
end if
if ((Root.LEFT).RIGHT < > NULL) then
add ((Root.LEFT). RIGHT);  // add child node in the Cluster
end if
Call: NCP(Root. RIGHT)
end if
end if
else if (Root. RIGHT < > NULL && Root. LEFT == NULL) then
if Root is not presented in Cluster then
makeCluster(Root, Root.RIGHT)
if ((Root. RIGHT). LEFT < > NULL) then
add (Root. RIGHT). LEFT  // add child node in the Cluster
end if
if ((Root.RIGHT). RIGHT < > NULL) then
Call: NCP((Root.RIGHT). RIGHT)
end if
end if
else if (Root.RIGHT == NULL && Root.LEFT < > NULL) then
if Root is not presented in Cluster then
makeCluster(Root, Root.LEFT)
if ((Root.LEFT). LEFT < > NULL) then
Call: NCP((Root. LEFT). LEFT)
end if
if ((Root.LEFT). RIGHT < > NULL) then
add ((Root. LEFT). RIGHT) ; // add child node in the Cluster
end if
end if
else
if Root is not presented in the Cluster then
makeCluster(Root)
end if
return
end if
END;
END.
```

3. A proposed forecasting model using Graph-based clustering and FTS

In this section, a forecasting model using fuzzy time series and graph - based clustering is introduced. The proposed model can be organized into following two main stages as shown in Figure 1: (1) Data partitioning stage (2) the stage of constructing of FTS forecasting model. The details of two these stages in the FTS forecasting model are explained according to steps as below. To handle these steps, all historical enrolments data [3] are utilized for illustrating forecasted process and which is shown in Figure 2. The six stages of forecasting model are described as follows:

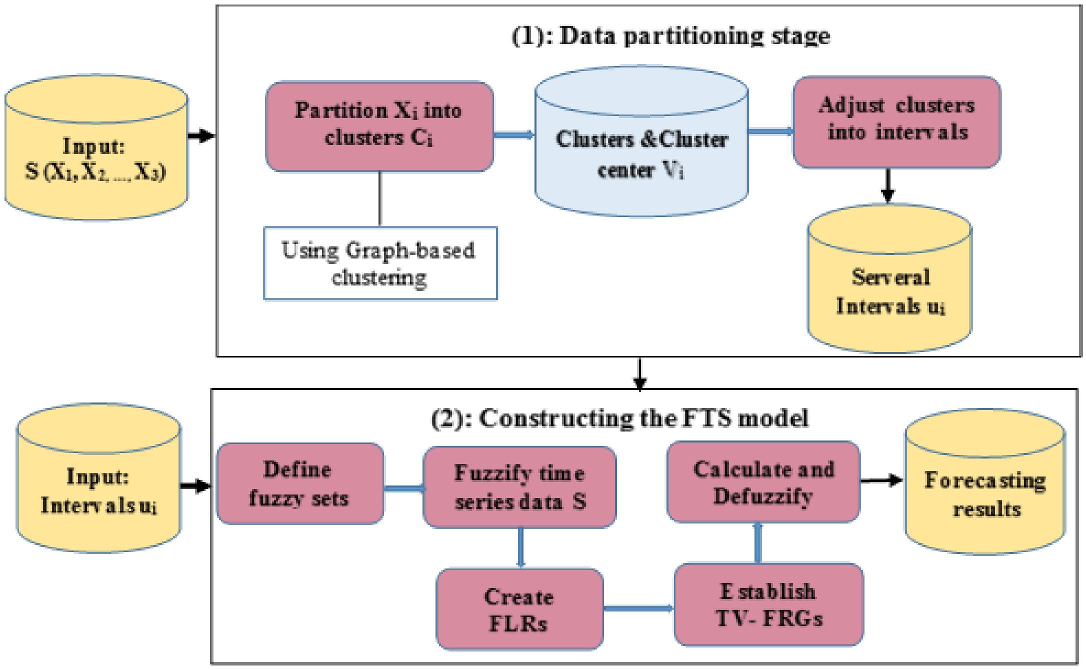

Figure. 1: Flowchart of the proposed forecasting model using FTS and Graph-based clustering

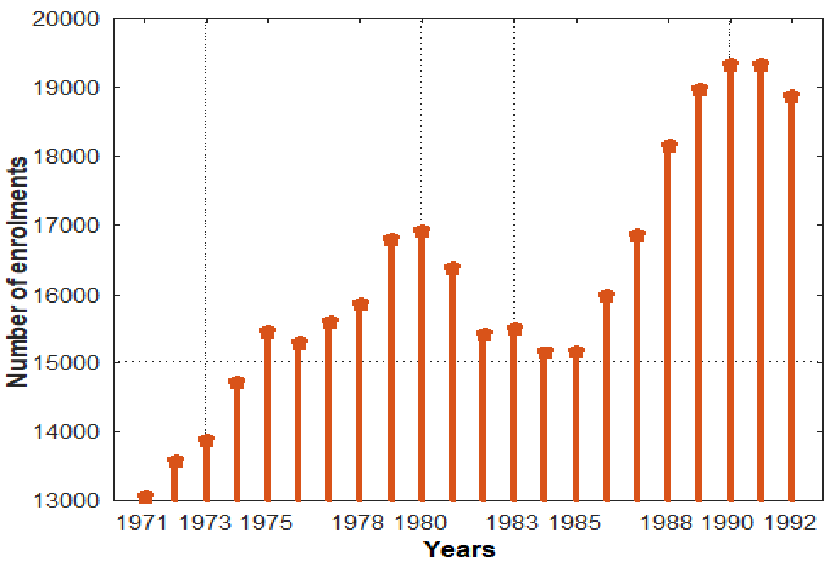

Figure 2: The historical data of enrolments

Step 1: Partitioning historical dataset of time series into intervals using graph-based clustering algorithm

This step, Graph-based clustering algorithm is applied to represent the time series dataset into clusters. After making the clusters, the clusters are adjusted into contiguous intervals with unequal- length.

Step 1.1: Apply **the Graph-based clustering algorithm** to partition data into **C** clusters.

To partition time series data into C clusters, four procedures of the Graph-based clustering algorithm in Section 2.2 are used. The brief results of these four procedures are explained as below

1) The Procedure of Finding Root Node (PFRN).
Input the car road accident data set of Alabama as:

S (13055*,* 13563*,* 13867*,* 14696*,* 15460*, . . . ,* 19328*,* 19337*,* 18876).

+ Calculate range *Rg* = MAX_{value} – MIN_{value} = 6282
+ Calculate standard deviation of the *time series as SD* = 1774.72

$$w = \frac{R_g}{SD*N} = 0.16$$

+ *Define UoD (U) of the S:*

U = [$\mathrm{MIN_{value}}$ − w, $\mathrm{MAX_{value}}$ + w] = [13054.84*,* 19337.16];

+ *Calculate midpoint of U as:*

Mid_u = ($\mathrm{MIN_{value}}$+ $\mathrm{MAX_{value}}$) / 2 =16196

+ *Assign the Mid_u as root node: root = Mid_u =16196*

2) Tree Creation Procedure (TCP) and Node Insertion Procedure (NIP).

For making the tree, from the input dataset *S* and *Root*. We utilize two procedures TCP and NIP to make tree and insert nodes into the tree. The results of these two procedures are shown in Figure 3.

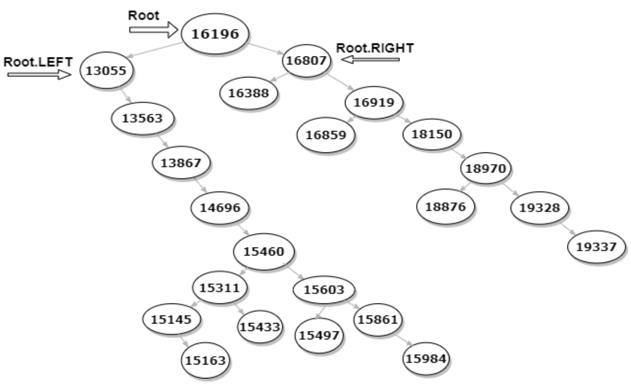

Figure 3: The tree represents the input data of time series based on two procedures TCP and NIP with root node of 16196

3) Make the clusters from the tree based on Procedure 4(NCP)

After obtaining the data tree shown in Figure 2, the procedure of making clusters is brief explained according to conditions as follows.

1. Initially, check that Root exists or not and check that Root has left (Root. LEFT) and right (Root. RIGHT)
2. If both children exist for each Root, then compute the difference between values of the Root and Root. RIGHT, and Root and Root. LEFT. Then make a cluster with corresponding child (either Root. LEFT or Root. RIGHT) and Root, which have the minimum difference.
3. If only one child exists for each Root, then make the cluster with either Root and Root. LEFT or Root and Root. RIGHT.
4. Repeat conditions 2-3, until all the value of the nodes in the tree are added to the clusters.

Based on Procedures of **the Graph-based clustering algorithm**, we achieve 10 clusters and their corresponding elements. Then, these clusters are sorted according to an ascending sequence of clustering centers, the final results are listed in Table 1.

Table 1: Clusters, their corresponding elements and center of clusters

Cluster	Corresponding Element	Center of clusters ($\mathrm{V_i}$)
C1	(16196, 16807, 16388)	13309
C2	(16919, 16859)	14281.5
C3	(18150, 18970, 18876)	15154
C4	(19328, 19337)	15372
C5	(13055, 13563)	15520
C6	(13867, 14696)	15904.5
C7	(15460, 15603, 15497)	16463.67
C8	(15861, 15984)	16889
C9	(15311, 15433)	18665.33
C10	(15145, 15163)	19332.5

Step 1.2: Adjust the clusters into intervals.

This step, then the upper bound and the lower bound of interval $\mathrm{I_i}$ can be defined by the minimum and maximum values of the corresponding clusters, respectively and the midpoint value of the intervals which are shown in Table 2.

Table 2: The completed results of intervals

No	Intervals	Mid_value
1	$\mathrm{I_1}$= [16196, 16807)	16292
2	$\mathrm{I_2}$= [16859, 16919)	16889
3	$\mathrm{I_3}$ = [18150, 18970)	18560
4	$\mathrm{I_4}$ = [19328, 19337)	19332.5
5	$\mathrm{I_5}$ = [13055, 13563)	13309
6	$\mathrm{I_6}$= [13867, 14696)	14281.5
7	$\mathrm{I_7}$= [15460, 15603)	15531.5
8	$\mathrm{I_8}$= [15861, 15984)	15922.5
9	$\mathrm{I_9}$= [15311, 15433)	15372
10	$\mathrm{I_{10}}$= [15145, 15163]	15154

Step 2: Determine linguistic terms for each of interval obtained in Step 1

The intervals obtained from Step 1 which are used to identify the linguistic terms. For ten intervals, we get ten linguistic values which is represented by fuzzy sets $\mathrm{A_i}$: $\{\mathrm{A_1}, \mathrm{A_2}, \mathrm{A_3}, \dots, \mathrm{A_6}, \mathrm{A_{10}}\}$ $(1 \le i \le 10)$ which are presented as: A_1={very very very few }, A_2={ very very few}, A_3={very few}, A_4 ={few }, A_5 ={ moderate}, A_6={many}, A_7={many many}, A_8={very many}, A_9={too many}, A_{10} ={too many many}.

$$\mathrm{A_i} = {a_{i1}}/{\mathrm{I_1}} + {a_{i2}}/{\mathrm{I_2}} + \ldots + {a_{ij}}/{\mathrm{I}_j} + \cdots + {a_{i10}}/{\mathrm{I_{10}}} \qquad (1)$$

Where $\mathrm{a_{ij}}$ ∈[0,1] is the membership grade of I_j belonging to $\mathrm{A_i}$, which is defined in (2), the symbol '+' denotes the set union operator and the symbol '/' denotes the membership of I_j which belongs to A_i.

$$a_{ij} = \begin{cases} 1 & j = i \\ 0.5 & j = i-1 \; or \; j = i+1 \\ 0 & otherwise \end{cases} \quad (2)$$

From (1), each fuzzy set contains 10 intervals, and each interval belongs to all fuzzy sets with different grade of membership values are presented in (2). For instance, I_1 corresponds to linguistic variables A_1 and A_2 with degree of membership values 1 and 0.5 respectively, and remaining fuzzy sets with membership grade is 0. The descriptions of remaining intervals, eg., I_2,..., I_{10}, can be explained in a similar way.

Step 3: Fuzzy all historical data

The way to fuzzify a historical data is to find the interval it belongs to and assign the corresponding linguistic value to it and finding out the degree of each data belonging to each A_i . If the maximum membership of the historical data is under A_i , then the fuzzified historical data is labeled as A_i.

For example, the historical enrolment of year 1973 is 13867 which falls within I_6 = (13867*,* 14696), so it belongs to interval I_6 Based on (1), Since the highest membership degree of I_6 occurs at A_6 is 1, the historical time variable F(1973) is fuzzified as A_6 . The same way for other years.

Step 4: Create all fuzzy logical relationships (FLRs)

Fuzzy relationships are identified from the fuzzified historical data. Support that the historical data of year i is A_i*,* then that of year i + 1 is A_k then these two consecutive fuzzified values are presented $A_i \rightarrow A_k$, where A_i is called the current state of the enrollment, and A_k is called the next state of the enrollment. This relation is called the first – order FLR.

Similarly, the fuzzy relationship can be constructed several consecutive fuzzified values, the relation between $\mathrm{F(t-n), F(t-n+1), \ldots, F(t-1)}$ and $\mathrm{F(t)}$ is denote by $\mathrm{F(t-n), F(t-n+1), \ldots, F(t-1) \rightarrow F(t)}$, it is called the n – order FLR.

Step 5: Establish all fuzzy relationship groups (FRGs)

In this study, we rely on the fuzzy relationship group [17] which are presented in Definition 5 to construct FRGs. To clarify this, we consider three first - order FLRs at three different times t-2, t-1 and t as follows: $A_i \rightarrow A_j$; $A_i \rightarrow A_k$ and $A_i \rightarrow A_j$, respectively. Suppose that we want to forecast the value of historical datum at time t-1, then fuzzy sets on the next state of FLRs having the same current state is considered to place into together G1 as $A_i \rightarrow A_j$, A_k, that is, only FLRs looking before time t-1 is clustered into a group together. The same way, if forecasting time t, the FLRs which have the same current state are grouped into a group G2 as $A_i \rightarrow A_j$, A_k, A_j.

Step 6: Defuzzify and calculate the forecasted values

Assume the fuzzified data of *F(t*-1) is A_j, then forecasted output value of *F(t)* is calculate by following principles.

Principle 1: If there exists a one-to-one relationship in the relationship group of A_j, say $A_j \rightarrow A_k$, and the highest degree of belongingness of A_k occurs at interval I_k, then the forecasted output of *F(t)* equals the midpoint of I_k.

Principle 2: If A_j is empty, i.e. $A_j \rightarrow \emptyset$, and the interval where A_j has the highest degree of belongingness is I_j, then the forecasted output equals the midpoint of I_j.

Principle 3: If there exists a one-to-many relationship in the relationship group of $A_j \rightarrow A_{i1}, A_{i2}, \ldots, A_{ip}$; then the forecasted output is computed as:

$$\boldsymbol{forcasted - value} = \frac{\boldsymbol{1 \times m_{i1} + 2 \times m_{i2} + \ldots + p \times m_{ip}}}{\boldsymbol{1 + 2 + \ldots + p}} \quad (3)$$

Where, $m_{i1}, m_{i2}, \ldots, m_{ip}$ are midpoint of $I_{i1}, I_{i1}, \ldots, I_{ip}$, respectively.

Step 7: Performance evaluation of the proposed model

The efficiency of the proposed forecasting model is evaluated using statistical indexes as the mean absolute percentage error (MAPE). The evaluation criterions are determined by the following equations:

$$MAPE = \frac{1}{n}\sum_{i=m}^{n}\left|\frac{F_i - R_i}{R_i}\right| * 100\% \quad (4)$$

Where, Ri and Fi note the actual and forecasted value at time i, respectively, n is the total number of years to be forecasted, m is the order of fuzzy logical relationship.

4. Experimental results and Analysis

In this paper, the proposed forecasting model has been applied to two time series, as enrolments data of University of Alabama [3] and numerical datasets of Gasonline Prices in Viet Nam. Before implementing the proposed forecasting model, the time series datasets are briefly described. Then, the simulated results and analyses related to these datasets are given,

4.1. Prepare time series data: a summary

1) The enrollment time series dataset

This time series data consists of 22 observation values between 1971 and 1992, S= { d_1, d_2, ..., d_{22}} as shown in Figure 2. This dataset has utilized to examined with the huge amount of reseach works which are presented in the articles [1-8, 16, 22, 35-39]. The obtained results among these research works are choosed for comparing with our proposed model. The UoD of enrolments time series is determined as $U = [d_{min}, d_{max}]$ = [13055, 19337]. In which, the d_{min}= min { d_1, d_2, ..., d_{22}} and d_{max}=max{d_1, d_2, ..., d_{22}}, respectively.

2) The Gas online prices time series dataset in Vietnam

The dataset consists of daily values of Gasonline price of E5 RON 92 of Viet Nam from January 30, 2020 to March 12, 2021, which consists of 26 data, as shown in Table 3, in which it taken from https://vnexpress.net/kinh-doanh/hang-hoa.

Table 3: The real data of Gasonline price of E5 RON 92

Date/ month	Real data	Date/ month	Real data
01/30/2020	19270	09/11/2020	14260
02/14/2020	18500	09/26/2020	14210
02/29/2020	18340	10/12/2020	14260
03/15/2020	16050	10/27/2020	14000
03/29/2020	11950	11/11/2020	13880
04/13/2020	11340	11/26/2020	14490
04/28/2020	10940	12/11/2020	15120
05/13/2020	11520	12/26/2020	15510
05/28/2020	12400	01/11/2021	15940
06/12/2020	13390	01/26/2021	16300
06/27/2020	14250	02/10/2021	16300
07/13/2020	14400	02/25/2021	17030
07/28/2020	14400	03/12/2021	17720

Source: https://vnexpress.net/kinh-doanh/hang-hoa

The minimum and the maximum of the time series are respectively d_{min} = 10940 and d_{max} = 19270. The universe of discourse of the time series can be determined, say *U* = [10940*,* 19270]

4.2. Experimental results

4.2.1. Experimental results for forecasting enrollments

In order to verify the performance of the forecasting model based on the first-order FTS under different number of intervals, the forecasting result obtained from the proposed model is compared with the ones of the current models [2, 3, 35-39]. A comparison with regards to MAPE value between the proposed model and the different forecasting models are given in Table 4. Considering the Table 4 the results show that the proposed model has the smallest forecasting errors with regards to MAPE value equal to 1.19% among all its counterparts.

Table 4: The forecasting results in our proposed model and compares it with other models in terms of MAPE

Year	Actual	Model [2]	Model [3]	Model[35]	Model[36]	Model[37]	Model[38]	Model[39]	Our model
1971	13055	-	-	-	-		-	-	
1972	13563	14000	14000	14025	15430	13944	14195	14242	13309
1973	13867	14000	14000	14568	15430	13944	14424	14242	13957.33
1974	14696	14000	14000	14568	15430	13944	14593	14242	14281.5
1975	15460	15500	15500	15654	15430	15328	15589	15474.3	15114.83
1976	15311	16000	16000	15654	15430	15753	15645	15474.3	15372
1977	15603	16000	16000	15654	15430	15753	15634	15474.3	15531.5
1978	15861	16000	16000	15654	15430	15753	16100	15474.3	15739
1979	16807	16000	16000	16197	16889	16279	16188	16146.5	16501.5
1980	16919	16813	16833	17283	16871	17270	17077	16988.3	16889
1981	16388	16813	16833	17283	16871	17270	17105	16988.3	16501.5
1982	15433	16789	16833	16197	15447	16279	16369	16146.5	15877.67
1983	15497	16000	16000	15654	15430	15753	15643	15474.3	15531.5
1984	15145	16000	16000	15654	15430	15753	15648	15474.3	15446.5
1985	15163	16000	16000	15654	15430	15753	15622	15474.3	15154
1986	15984	16000	16000	15654	15430	15753	15623	15474.3	15666.33
1987	16859	16000	16000	16197	16889	16279	16231	16146.5	16759.83
1988	18150	16813	16833	17283	16871	17270	17090	16988.3	17873.83
1989	18970	19000	19000	18369	19333	19466	18325	19144	18560
1990	19328	19000	19000	19454	19333	18933	19000	19144	19075
1991	19337	19000	19000	19454	19333	18933	19000	19144	19332.5
1992	18876		19000		19333	18933	19000	19144	18817.5
MAPE	-	3.22%	3.11%	2.67%	2.75%	2.68%	2.66%	2.40%	1.19%

To be clearly imagined, Figure 4 describes the trend in terms of MAPE between our model and the previous models for different intervals. Viewing these curves, it is clearly seen that forecasting accuracy of our proposed model is more accurate than those of compared models under the first-order FLR.

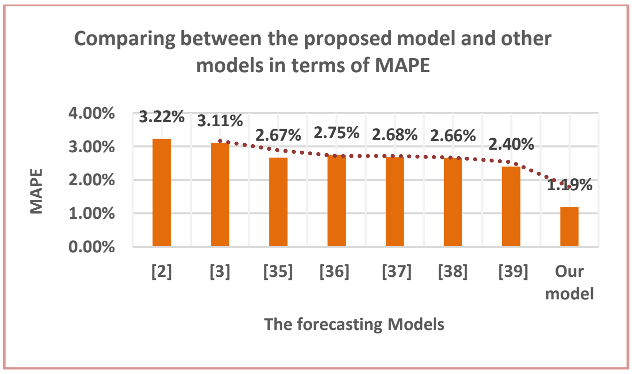

Figure 4: The graph represents the forecasting accuracy of the proposed model and other models

4.2.1. Experimental results for forecasting the Gasonline price of E5 RON 92

In this section, we apply the proposed model to forecast the Gasonline price of Viet Nam with the whole historical data the period from 1/30/2020 to 03/12/2021 is shown in Table 3. We implement the proposed model under different number of orders and kept number of intervals equal to 9 which is obtained by Graph-based clustering algorithm. The forecasted accuracy of the proposed model is estimated by using the MAPE function (4). The forecasted results of proposed model are depicted in Figure 5.

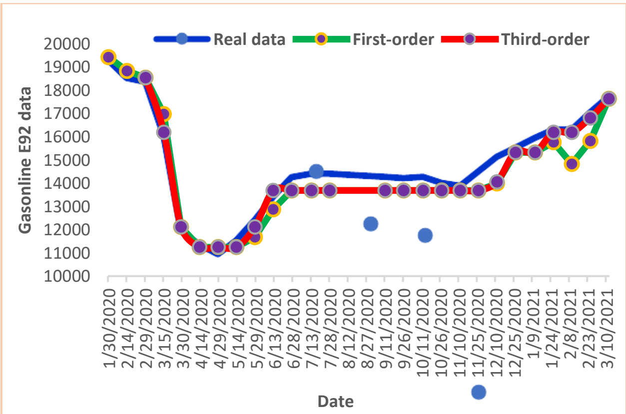

Figure 5: The curves represent the real values and forecasted values of the proposed model under different orders

Conclusion

In this study, the novel partitioning method based on the graph clustering is proposed for improving forecasting performance in the different application. The main contributions of this paper are illustrated in the following. Firstly, a new algorithm for finding the different interval lengths based on a graph - based clustering algorithm is proposed. Then, the various fuzzy sets based on these evolved intervals are defined and the historical data are fuzzified. Based on these fuzzified values, we derive the FLRs and group the weighted fuzzy relationship according to their chronological order. From obtained results in Table 4 and Figures 3, 4 clearly show that using unequal-sized partitioning can produce better forecasting accuracy than the equal-sized intervals and the forecasting effectiveness of the proposed model outperforms previous forecasting models.

Acknowledgement

This work was supported in part by the Science Council of Thai Nguyen University of Technology (TNUT). The authors are grateful to the anonymous reviewers for their comments which were very helpful in improving the quality of the research.

References

[1] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series – Part I," Fuzzy Sets and Systems, vol. 54, no. 1, pp. 1–9, 1993.
[2] L. A. Zadeh, "Fuzzy sets," Information and Control, vol. 8, no. 3, pp. 338–353, 1965.
[3] S. M. Chen, "Forecasting enrollments based on fuzzy time series," Fuzzy Sets and Systems, vol. 81, pp. 311–319, 1996.
[4] Q. Song and B. S. Chissom, "Fuzzy time series and its models," Fuzzy Sets and Systems, vol. 54, no. 1, pp. 1–9, 1993.
[5] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series – Part II," Fuzzy Sets and Systems, vol. 62, no. 1, pp. 1–8, 1994.
[6] H.-T. Liu and M.-L. Wei, "An improved fuzzy forecasting method for seasonal time series," Expert Systems with Applications, vol. 37, no. 9, pp. 6310–6318, 2010.
[7] S.-M. Chen and K. Tanuwijaya, "Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques," Expert Systems with Applications, vol. 38, no. 8, pp. 10 594–10 605, 2011.
[8] Y.-L. Huang, S.-J. Horng, T.-W. Kao, R.-S. Run, J.-L. Lai, R.-J. Chen, I.-H. Kuo, and M. K. Khan, "An improved forecasting model based on the weighted fuzzy relationship matrix combined with a PSO adaptation for enrollments," International Journal of Innovative Computing, Information and Control, vol. 7, no. 7 (A), pp. 4027–4045, 2011.
[9] L. Wang, X. Liu, and W. Pedrycz, "Effective intervals determined by information granules to improve forecasting in fuzzy time series," Expert Systems with Applications, vol. 40, no. 14, pp. 5673–5679, 2013.
[10] W. Lu, X. Chen, W. Pedrycz, X. Liu, and J. Yang, "Using interval information granules to improve forecasting in fuzzy time series," International Journal of Approximate Reasoning, vol. 57, pp. 1–18, 2015.
[11] J. R. Hwang, S. M. Chen, and C. H. Lee, "Handling forecasting problems using fuzzy time series," Fuzzy Sets and Systems, vol. 100, pp. 217–228, 1998.
[12] K.-H. Huarng and T. H.-K. Yu, "Modeling fuzzy time series with multiple observations," International Journal of Innovative Computing, Information and Control, vol. 8, no.10(B), pp. 7415–7426, 2012. [13] P. Singh and B. Borah, "An efficient time series forecasting model based on fuzzy time series," Engineering Applications of Artificial Intelligence, vol. 26, pp. 2443–2457, 2013.
[14] T.-L. Chen, C.-H. Cheng, and H. J. Teoh, "Fuzzy time-series based on Fibonacci sequence for stock price forecasting," Physica A: Statistical Mechanics and its Applications, vol. 380, pp. 377–390, 2007.

[15] H.-K. Yu, "Weighted fuzzy time series models for TAIEX forecasting," Physica A: Statistical Mechanics and its Applications, vol. 349, no. 3–4, pp. 609–624, 2005.
[16] S.-M. Chen, "Forecasting enrollments based on high-order fuzzy time series," Cybernetics and Systems: An International Journal, vol. 33, no. 1, pp. 1–16, 2002.
[17] T.-L. Chen, C.-H. Cheng, and H.-J. Teoh, "High-order fuzzy time-series based on multiperiod adaptation model for forecasting stock markets," Physica A: Statistical Mechanics and its Applications, vol. 387, no. 4, pp. 876–888, 2008.
[18] S. R. Singh, "A computational method of forecasting based on high-order fuzzy time series," Expert Systems with Applications, vol. 36, no. 7, pp. 10 551–10 559, 2009.
[19] S.-M. Chen and K. Tanuwijaya, "Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques," Expert Systems with Applications, vol. 38, no. 12, pp. 15 425–15 437, 2011.
[20] S. S. Gangwar and S. Kumar, "Partitions based computational method for high-order fuzzy time series forecasting," Expert Systems with Applications, vol. 39, no. 15, pp. 12 158–12 164, 2012.
[21] K. Huarng, "Heuristic models of fuzzy time series for forecasting," Fuzzy Sets and Systems, vol. 123, pp. 369–386, 2001.
[22] S. R. Singh, "A robust method of forecasting based on fuzzy time series," Applied Mathematics and Computation, vol. 188, no. 1, pp. 472–484, 2007.
[23] S.-T. Li and Y.-C. Cheng, "Deterministic fuzzy time series model for forecasting enrollments," Computers and Mathematics with Applications, vol. 53, no. 12, pp. 1904–1920, 2007.
[24] H. S. Lee and M. T. Chou, "Fuzzy forecasting based on fuzzy time series," International Journal of Computer Mathematics, vol. 81, no. 7, pp. 781–789, 2004.
[25] P. Singh and B. Borah, "An effective neural network and fuzzy time series-based hybridized model to handle forecasting problems of two factors," Knowledge and Information Systems, vol. 38, no. 3, pp. 669–690, 2012.
[26] S. M. Chen and J. R. Hwang, "Temperature prediction using fuzzy time series," IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 30, pp. 263– 275, 2000.
[27] Z.-Y. Lee, "Method of bilaterally bounded to solution blasius equation using particle swarm optimization," Applied Mathematics and Computation, vol. 179, no. 2, pp. 779– 786, 2006.
[28] L.-W. Lee, L.-H. Wang, and S.-M. Chen, "Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms," Expert Systems with Applications, vol. 33, no. 3, pp. 539–550, 2007.
[29] L.-W. Lee, L.-H. Wang, and S.-M. Chen, "Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationships and genetic simulated annealing techniques," Expert Systems with Applications, vol. 34, no. 1, pp. 328–336, 2008.
[30] Y. C. Chang and S. M. Chen, "Temperature prediction based on fuzzy clustering and fuzzy rules interpolation techniques," in Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics, San Antonio, TX, USA, October 2009, pp. 3444–3449.
[31] N.-Y. Wang and S.-M. Chen, "Temperature prediction and TAIFEX forecasting based on automatic clustering techniques and two-factors high-order fuzzy time series," Expert Systems with Applications, vol. 36, no. 2, Part 1, pp. 2143–2154, 2009.
[32] P. Singh and B. Borah, "Indian summer monsoon rainfall prediction using artificial neural network," Stochastic Environmental Research and Risk Assessment, vol. 27, no. 7, pp. 1585–1599, 2013.
[33] M.-Y. Chen, "A high-order fuzzy time series forecasting model for internet stock trading," Future Generation Computer Systems, vol. 37, pp. 461–467, 2014.
[34] Nghiem Van Tinh, Nguyen Cong Dieu, Handling Forecasting Problems Based on Combining High-Order Time-Variant Fuzzy Relationship Groups and Particle Swam Optimization Technique", International Journal of Computational Intelligence and Applications, Vol 17(2), (2018), 1-19.
[35] H. S. Lee and M. T. Chou, "Fuzzy forecasting based on fuzzy time series," International Journal of Computer Mathematics, vol. 81, no. 7, pp. 781–789, 2004.
[36] C. Cheng, J. Chang, and C. Yeh, "Entropy-based and trapezoid fuzzification-based fuzzy time series approaches for forecasting IT project cost," *Technological Forecasting and Social Change*, vol. 73, pp. 524–542, 2006.
[37] L. Wang, X. Liu, W. Pedrycz, Y. Shao, Determination of temporal information granules to improve forecasting in fuzzy time series, Expert Syst. Appl. 41 (6) (2014) 3134–3142, http://dx.doi.org/10.1016/j.eswa.2013.10.046.
[38] W. Qiu, X. Liu, and H. Li, "A generalized method for forecasting based on fuzzy time series," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10 446–10 453, 2011.
[39] C. H. Cheng, G. W. Cheng, and J. W. Wang, "Multi-attribute fuzzy time series method based on fuzzy clustering," *Expert Systems with Applications*, vol. 34, pp. 1235–1242, 2008.