ISSN (e): 2250 – 3005 || Volume, 14 || Issue, 2|| Mar. - Apr. – 2024 ||
International Journal of Computational Engineering Research (IJCER)

Enhancing Forecasting Accuracy in Fuzzy Time Series Model Utilizing Graph-Based Clustering

Nghiem Van Tinh[1,], Pham Quang Hieu[1]
[1]Faculty of Electronics - Thai Nguyen University of Technology, Thai Nguyen, Vietnam
[*]Corresponding Author: Nghiem Van Tinh

ABSTRACT
Over the years, many fuzzy time series (FTS) forecasting models have emerged to tackle complex and incomplete problems. However, the accuracy of these models varies depending on the specific problem and dataset used. Despite claims of being better than traditional statistical and single machine learning-based models, improving forecasting accuracy remains a tough challenge. In FTS models, the lengths of intervals and groups of fuzzy relationships are seen as key factors affecting accuracy. This study introduces an FTS forecasting model based on graph-based clustering. The clustering algorithm used during the fuzzification stage allows for different interval lengths to be determined. The proposed model is tested on two numerical datasets: enrollment data from the University of Alabama and Gas prices RON95 datasets in Vietnam. Comparisons of forecasting results between the proposed model and others are made for enrollment forecasts at the University of Alabama. The results show that the proposed model consistently achieves higher forecasting accuracy across all levels of fuzzy relationships compared to existing models.
.
KEYWORDS: *Forecasting, FTS, fuzzy logical relationss, graph - based clustering, enrollments*

Date of Submission: 02-04-2024 Date of acceptance: 13-04-2024

I. INTRODUCTION

Forecasting daily events like temperature, stock market, enrollments and car road accidents poses a significant challenge in forecasting. Achieving perfect accuracy may not be feasible, but minimizing errors is crucial. Traditional models like Autoregressive , Moving Average , and Autoregressive Integrated Moving Average rely on linear assumptions and extensive historical data. Conversely, Fuzzy Time Series (FTS) models proposed by Song and Chissom [1, 2], offer flexibility with fuzzy relational matrices but struggle with interval determination and computational complexity. To address these issues, Chen's [3] first-order FTS model employs simpler arithmetic operations, leading to improved prediction accuracy. Subsequent research has explored enhancements including equal and different interval lengths [4-7], refined fuzzy relationship groups [8-14], and defuzzification processes [15-19]. Recently, numerous researchers have integrated intelligent computational techniques with various FTS models to tackle intricate forecasting challenges. For instance, Lee et al. [20] explored a two – factor high-order FTS model for temperature prediction and TAIFEX. They also employed an annealing technique [21] to optimize division lengths for enhanced forecasting accuracy. Additionally, Chen and Chung [22] utilized a genetic algorithm (GA) to optimize intervals within the universal of discourse, introducing both first-order and high-order forecasting models to predict Alabama university enrollments. Another approach, Chen and Wang [23] employs involves using automatic clustering techniques to build fuzzy time series forecasting models for temperature and TAIFEX (Taiwan Futures Exchange) problems. Presently, the utilization of particle swarm optimization (PSO) to select optimal intervals in fuzzy time series forecasting models has garnered considerable attention. Research demonstrates that PSO-based interval selection significantly enhances forecasting model performance [24, 25]. For instance, Kuo et al. [13] proposed a novel forecasting model by integrating PSO with FTS to enhance prediction accuracy and devised a new defuzzification rule for TAIFEX prediction. Similarly, studies [23, 26] introduced two-factor high-order FTS models for forecasting the Taiwan stock market and TAIFEX, utilizing PSO for interval optimization. Furthermore, Park et al. [27] proposed a PSO-enhanced two-factor high-order FTS model for more accurate forecasting results. Huarng et al. [19] introduced a hybrid predictive model incorporating PSO to rectify output prediction rules for university admissions. Research work [16] proposes a novel probabilistic intuitionistic FTS forecasting model using support vector machine to address both uncertainty and non-determinism associated with real world time series data. Cheng et al. [28] proposed an FTS model for TAIFEX prediction, employing PSO for interval length determination and the K-means algorithm for fuzzy set indexing. Various techniques such as raw clustering [9], automatic clustering [10], and fuzzy C-Mean clustering [25] have been introduced to determine interval length in recent studies.

Analyzing the works mentioned above reveals that determining the appropriate interval length, establishing fuzzy relationships, and formulating output prediction rules pose significant challenges and greatly impact predictive accuracy. Moreover, incorporating observed factors beyond the main forecasted factor is essential for enhancing predictive efficiency. Despite notable progress in determining interval length and exploiting output prediction rules, these challenges continue to captivate researchers' attention. In an effort to enhance the predictive efficiency of FTS forecasting models, this study proposes a novel forecasting model utilizing a graph-based clustering technique to determine interval lengths using dataset of Gas prices RON95 in Vietnam. Initially, we propose a new algorithm for optimal interval length determination using a graph-based clustering approach. Subsequently, we define fuzzy sets based on these intervals and fuzzify historical data. Based on these fuzzified values, fuzzy relationships are derived, followed by the formation of fuzzy relationship groups (FRGs) according to chronological order. Finally, these FRGs are utilized to derive forecasting results employing a weighted defuzzification method.

The subsequent parts of this paper follow this organization: Section 2 elucidates fundamental definitions of fuzzy time series and algorithms. In Section 3, we introduce a forecasting model that combines FTS with a Graph-based clustering algorithm. Section 4 evaluates the models' performance and compares the results with those obtained from other models. Lastly, Section 5 delivers concluding remarks.

II. THE FUNDAMENTAL THEORIES

In this section, we briefly introduce general knowledge related to FTS which is proposed in [1-3] and improved by research work in [4].

2.1. Fuzzy time series

The concepts of FTS were defined by Song and Chissom [2, 3], in which the historical data are given in the form of fuzzy sets [1]. Assume that $Y(t)$ $(t = .., 0, 1, 2 ..)$ a real subset R ($Y(t) \subseteq$ R), regarded as the UoD on which the fuzzy sets $f_i(t)$ $(i = 1,2 ...)$ are defined. If $F(t)$ including the collection of $f_1(t), f_2(t), ...$, then $F(t)$ is namely an FTS which is defined on $Y(t)$.

If there exists fuzzy relationship (FR) between $F(t-1)$ and $F(t)$, namely $R(t-1,t)$, such that they can be expressed as $F(t) = F(t-1) * R(t-1,t)$ or $F(t-1) \rightarrow F(t)$; Where $R(t-1,t)$ is the first- order fuzzy relationship between F(t) and $F(t-1)$ and "$*$" represents the max–min composition operator. Here $F(t)$ and $F(t-1)$ are fuzzy sets. If, let A_i = F(t) and A_j = F(t-1), the relationship between $F(t)$ and $F(t-1)$ is replaced by $A_i \rightarrow A_j$, where A_i and A_j are called the current state and the next state of fuzzy relationship, respectively.

Let F(t) be a fuzzy time series. If F(t) is derived by more fuzzy sets $F(t-1), F(t-2), ..., F(t-m+1), F(t-m)$, then fuzzy relationship between them can be represented as $F(t-m), ..., F(t-2), F(t-1) \rightarrow F(t)$. This relationship is called the m - order FTS model [3]

The fuzzy logical relationships having the same left- hand side can be further grouped into a Fuzzy relationship group [4]. 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 : $A_i \rightarrow A_{k1}, A_{k2}, \ldots, A_{km}$.

In order to generate the forecasting rules, the approach involves building fuzzy relationships through time-variant grouping, also known as censusing [25]. This entails computing the fuzzy relationships established from the training dataset based on identical left-hand sides and right-hand sides at the time of forecasting relative to previous instances.

2.2. Graph based clustering algorithm (GBC)

Graph-based clustering algorithms have demonstrated a remarkable ability to produce results that align closely with human intuition [29]. A defining feature shared in graph-based clustering methods is the utilization of a graph constructed from the dataset during the clustering process [30]. In these methods, data entities are represented as nodes within a graph, with connections established between related entities. Clusters are formed when a group of entities is interconnected but lacks connectivity to entities outside the group. Building upon these principles, our research introduces a novel data clustering approach [31], wherein the dataset is represented as a tree structure, and clusters are automatically generated without requiring the user to pre-specify the number of clusters. Specifically, the graph-based clustering method can be outlined in four distinct procedures as follows:

(1) Root node location procedure (RNLP). This technique identifies the root node based on the provided data.
(2) Node insertion procedure (NIP). This technique inserts one element of the dataset and root node and places the elements in the proper position.
(3) Tree making procedure (TMP). This procedure displays the tree from the provided data set and the root node.
(4) Clustering procedure (CP) based on nodes in the tree. This process makes logical node clustering using the tree that the TMP generated as input.

III. A FUZZY TIME SERIES FORECASTING MODEL USING GRAPH- BASE CLUSTERING

The aim of this section is to introduce a hybrid FTS forecasting model that integrates Graph-based clustering and FTS. The proposed forecasting model framework comprises six steps, as illustrated in Figure 1. To address these steps, the datasets of Gas prices RON95 in Vietnam from 02/10/2023 to 28/03/2024 are depicted in Figure 2, which serve to illustrate the forecasting process. The specifics of the steps in the proposed model are elaborated as follows:

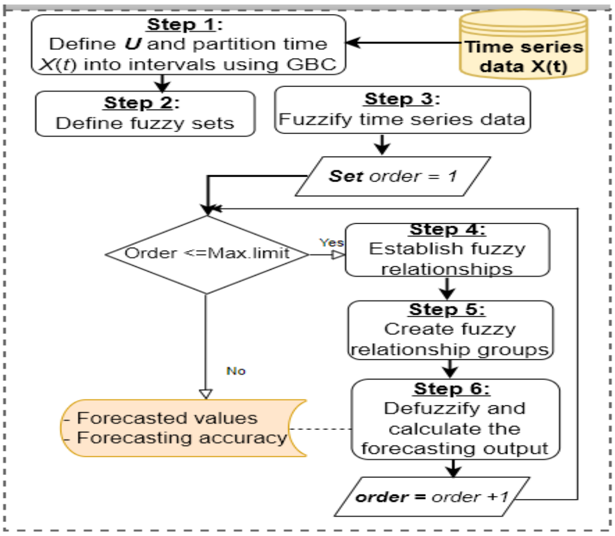

Figure 1: The flowchart illustrating the proposed forecasting model employing Graph-based clustering

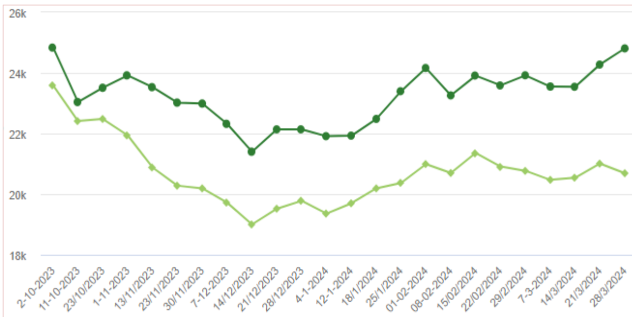

Figure 2: The datasets of Gas prices RON95 in Vietnam which was collected from https://vnexpress.net/kinhdoanh/hang-hoa

Step 1: Partitioning dataset into intervals using graph-based clustering

This step applies graph-based clustering algorithm in Section 2.2 to partition historical dataset into clusters, and then, adjust them into intervals with unequal-size. The calculation is described according to sub-steps as follows:

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

To partition time series $X(t)$ into C clusters, four procedures of the graph-based clustering algorithm in Section 2.2 are used in this step. The brief results of these four procedures are explained as below:

1) Root node location procedure.

Input the dataset of of Gas prices RON95 as: $X(t)$ = (24840, 23040,..., 24280, 24810) with (02/10/2023 ≤ t ≤ 28/03/2024).

Calculate range Rg = $MAX_{value} - MIN_{value}$ = 3440

Calculate standard deviation of the time series as *SD* = 925.07: $w = \frac{R_g}{SD*N} = 0.15$

Define universe of discourse (U) of the S: U = [$\text{Min}_{\text{value}}$ – w, $\text{Max}_{\text{value}}$ + w] = [21399.85, 24840.15];

Calculate midpoint of U: Mid_{u} = ($\text{Min}_{\text{value}}$+ $\text{Max}_{\text{value}}$) / 2 = 23120

Assign the Mid_{u} as root node: Root = Mid_{u} =23120

2) Node insertion procedure and Tree making procedure.

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

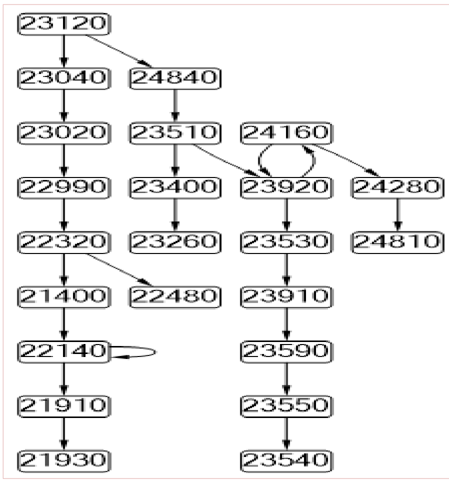

Figure 3: The tree represents the input data of Gas prices RON95 time series

3) Make the clusters based on Procedure 4 (CP)

After creating the data tree as shown in Figure 3, 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 existing 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.

From Procedures above, we achieve 7 clusters and their corresponding cluster centers. Then, these clusters are sorted according to an ascending sequence of clustering centers, the final results are listed in Table 1.

Table 1: The completed clusters from the Ron95 dataset

Number of clusters	Clusters C_i
1	(21400, 22140, 21910)
2	(22140)
---	---
6	(23400, 23260, 23920)
7	(24840,23510)

Step 1.2: *Adjust the clusters into intervals.*

Based on the obtained clusters, we adjust them into contiguous intervals according to the principles [25]. This step, the clusters in Table 1 are altered into intervals in accordance with cluster centres. Assume that V_{i+1} is the adjacent clustering centre next to V_i and each cluster C_I is assigned as an interval u_i , then the upper bound value of interval u_i ($Interval\_UB_i$) and the lower bound value of interval u_{i+1} ($Interval\_LB_{i+1}$) can be computed as follows:

$$Interval\_UB_i = \frac{V_i + V_{i+1}}{2} \quad (1)$$

$$Interval\_LB_{i+1} = Interval\_UB_i \quad (2)$$

Because of lacking of the lower bound value for the first interval and lacking of the upper bound value for the last interval, the lower and upper bounds of these intervals are formed from the minimum element of the first cluster and the maximum element of the last cluster, respectively.

Compute midpoint value of the interval interval_I as follows:

$$\text{Mid\_point}_\text{I} = \frac{\text{Interval\_LB}_\text{i} + \text{Interval\_UB}_\text{i}}{2} \tag{3}$$

After applying the conditions above, we obtain 7 intervals corresponding to the clusters in Table 1, called u_I $(1 \leq \text{i} \leq 7)$ and the midpoint values of these intervals are shown in Table 2.

Table 2: The result of intervals and it's midpoints

No	Intervals (u_i)	Midpoint
1	[21400 – 21978.35)	21689.2
2	[21978.35, 22270)	22124.2
---	---	---
6	[23303.35, 23850.85)	23577.1
7	[23850.85, 24840]	24345.4

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

Each linguistic term can be defined by intervals that the historical time series data is distributed among these intervals. For ten intervals in step 1, we obtain 7 linguistic values of linguistic variables "RON95" which can be represented by fuzzy sets A_I, eg, $\{A_1, A_2, \dots, A_6, A_7\}$, respectively and calculated as follows:

$$\text{A}_\text{I} = {}^{a_{i1}}/_{\text{u}_1} + {}^{a_{i2}}/_{\text{u}_2} + \ldots + {}^{a_{ij}}/_{\text{u}_j} + \cdots + {}^{a_{i7}}/_{\text{u}_7} \tag{4}$$

Where, the values $\text{a}_{\text{ij}} \in [0,1]$ indicates the grade of membership of u_j in fuzzy set A_i . The degree of each data is determined according to their membership grade to the fuzzy sets and which is defined in (5). Here, the symbol '+' denotes the set union operator and the symbol '/' denotes the membership of u_j which belongs to A_i. The value of a_{ij} is defined as follows:

$$a_{ij} = \begin{cases} 1 & j = i \\ 0.5 & j = i-1 \; or \; j = i+1 \\ 0 & others \end{cases} \tag{5}$$

Step 3: Fuzzy all historical time series data

Each interval obtained in Step 1 can cover one or more historical data value of time series. In order to all historical time series, the common way is to convert historical data which belongs to the interval U into fuzzy sets. If the maximum membership value of fuzzy set A_i occurs at u_i, then the fuzzified historical value is considered as A_i**.** For example, the RON95 data on day 02/10/2023 equal to 24840 belongs to the interval u_7 = [23850.85, 24840] and the highest membership value of fuzzy set A_7 occurs at u_7 So, it is fuzzified into A_7. The similar way for next years, we complete the results of fuzzification of enrolments data for all years, as listed in Table 3.

Table 3: The complete fuzzified results

Day	Actual data	Fuzzy sets
02/10/2023	24840	A_7
11/10/2023	23040	A_4
-----	-----	----
21/03/2024	24280	A_7
28/03/2024	24810	A_7

Step 4: Create all m^{th}- order FRs between the fuzzified data values. ($m \geq 1$).

After converting data values of time series into fuzzy sets, the m^{th}- order FRs is created between two or many consecutive fuzzified values in time series. For establishing of these relationships, we need to find any relationship which has the type $F(t-m), F(t-m+1), \dots, F(t-1) \rightarrow F(t)$, where, the left - hand side of FR is called the current state and the right - hand side of FR is called next state, respectively. Then, the m^{th} - order FR is replaced by relation in accordance with the corresponding fuzzy sets as: $A_{im}, A_{i(m-1)}, \dots, A_{i2}, A_{i1} \rightarrow A_k$. For example, with $m = 1$. From Table 4, it can be seen that the fuzzified historical data of time series on the day $t-1$ of 02/10/2023 and t of 11/10/2023 are fuzzy sets A_7 and A_4 , respectively. The structure of the first - order FRs is created by two consecutive fuzzy sets as $A_7 \rightarrow A_4$, we have achieved the 1st-order FRs for the all fuzzified data values, which are presented in column 4 of Table 4.

Where, the linguistic value of F(28/03/2024) on the right - hand side of the last relationship is denoted by symbol '#' which is used to represent the unknown linguistic value.

Table 4: The complete the 1st – order fuzzy relationships

Day	No	Fuzzy set	1st – order FRs
02/10/2023		A_7	
11/10/2023	1	A_4	$A_7 \rightarrow A_4$
23/10/2023	2	A_6	$A_4 \rightarrow A_6$
	--	----	----------
21/03/2024	22	A_7	$A_6 \rightarrow A_7, A_4, A_7, A_7, A_6, A_7$
28/03/2024	23	A_7	$A_7 \rightarrow A_4, A_6, A_5, A_6, A_6, A_7$

Step 5: Generate all m – order FRGs

In this study, we apply the concept of time - variant fuzzy relationship group [25] to create FRGs. Based on the current state of the FRs in Table 4, the FRs can be grouped into a FRG by considering the history of appearance of the fuzzy sets on the next state of the FRs, and called time variant-FRGs. From this viewpoint, we obtain all 1st - order FRGs, which are shown in Table 5. Where, there are 30 groups in training phase and one group in testing phase.

Table 5: The complete the first–order fuzzy relationship groups

No	1st – order FRs	1st – order FRGS
G1	$A_5 \rightarrow A_5$	$A_5 \rightarrow A_5$
G2	$A_5 \rightarrow A_6$	$A_5 \rightarrow A_6$
----	----------	--------------
G22	$A_6 \rightarrow A_7$	$A_6 \rightarrow A_7, A_4, A_7, A_7, A_6, A_7$
G23	$A_7 \rightarrow A_7$	$A_7 \rightarrow A_4, A_6, A_5, A_6, A_6, A_7$
G24	$A_7 \rightarrow \#$	$A_7 \rightarrow \#$

Step 6: Defuzzify and compute the forecasting output values

In order to defuzzify the fuzzified data values, the defuzzified principle in article [14] is presented to compute the forecasted value for all 1st – order and high – order FRGs in training phase. Next, we use a defuzzified principle [26] for computing with the unknown linguistic value in testing phase. The forecasting principles is presented as follows:

Rule 1: Calculate the forecasting value with known linguistic values

To obtain the forecasting output results of proposed model from the fuzzy relation groups. Based on [14], we create all forecast outputs for fuzzy logical relationship groups based on fuzzy sets on the right-hand or next state within the same group. For each group in Table 5, we divide each corresponding interval of each next state into q sub-regions with equal size, and create a forecasted value for each group according to equation (6).

$$\text{forecasted}_{\text{output}} = \frac{1}{n}\sum_{j=1}^{n}\frac{(m_{kj}+\text{subm}_{kj})}{2} \tag{6}$$

Where, *FV* is forecasted value at time *t*, *n* is the sum of fuzzy sets on the next state of FRG.

- n is the total number of next states or the total number of fuzzy sets on the right-hand side within the same group.
- m_{kj} $(1 \leq j \leq n)$ is the midpoint of interval u_{kj} corresponding to j-th fuzzy set on the right-hand side where the highest level of fuzzy set A_{kj} takes place in these intervals, u_{kj}.
- subm_{kj} is the midpoint of one of q sub-regions corresponding to j-th fuzzy set on the right-hand side where the highest level of A_{kj} takes place in this interval.

Rule 2: Calculate the forecasting value with unknown linguistic values

In the testing phase, we calculate forecasting value for the group of fuzzy relationship which has the unknown linguistic value appearing in the next state. Assume that there is the mth-order fuzzy relationship group whose next state is #, shown as follows: $A_{i\,\text{m}}, A_{i\,\text{m}-1}, \ldots, A_{i\,1} \rightarrow \#$.

Where the symbol ‘‘#” denotes an unknown value, then the forecasted value of year i is identified according to [26] as follows:

$$\text{FV} = m_{i1} + \frac{\sum_{k=2}^{m} m_{i(k-1)} - m_{ik}}{2^{k-1}} \tag{7}$$

where, $m_{i\,1}, m_{i2}, \ldots, m_{ik}$ is midpoints of u_{i1}, u_{i2}, . . ., and u_{ik} $(2 \leq k \leq m)$, respectively.

Based on two forecasting principles above, we complete forecasting results for Gas prices RON95 in Vietnam from 02/10/2023 to 28/03/2024 based on first-order FRGs under seven intervals, which are listed in Table 6.

Table 6: The complete forecasted results based on the 1st - order FRGs

Day	RON 95 data	Fuzzy sets	Forecasted values
02/10/2023	24840	A_7	Not forecasted
11/10/2023	23040	A_4	22936.2
-----	-----	----	-------
21/03/2024	24280	A_7	23894.4
28/03/2024	24810	A_7	23564.3

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1. Datasets and evaluation criterion

In this research, two illustrative datasets including the historical data of enrollments at the University of Alabama [3] from 1971 to 1992 and the datasets of Gas prices RON95 in Vietnam which was collected from https://vnexpress.net/kinhdoanh/hang-hoa are selected to compare the proposed model with some other forecasting models. To confirm the effectiveness of proposed forecasting model on two theses datasets, mean square error (MSE) and mean absolute percentage error (MAPE) are employed as an evaluation criterion in term of the forecasted accuracy. The MSE and MAPE can be calculated as follows:

$$MSE = \frac{1}{n}\sum_{i=m}^{n}(F_i - R_i)^2 \tag{8}$$

$$MAPE = \frac{1}{n}\sum_{i=m}^{n}\left|\frac{F_i - R_i}{R_i}\right| * 100\% \tag{9}$$

Where *R*, *F* represent the real value and forecasting value at year i, respectively; *n* is the total number of forecasted data, *m* means the order of the fuzzy relationships.

4.2. Forecasting the enrollments of the University of Alabama

In this section, we evaluate proposed forecasting model in education domain on enrolments data of University of Alabama and compared the obtained results with previous prediction models [3, 31-35] to demonstrate the performance of our method. The obtained forecasting results from the proposed model which are shown in Table 7. The results in Table 7 show that the proposed model has the MSE(8) value of 57294.53 which is the smallest among all the models compared with number of intervals equal to 7. This can be seen that the proposed model gives a very positive predictive effect on the enrollments problem of the University of Alabama. The trend in forecasting of enrollments by first-order fuzzy time series model in comparison to the actual enrollments and with other existing models can be visualized in Fig. 4. From forecasted values in Table 7 and Fig. 4, it is also found that integration of the GBC technique with the fuzzy time series model reduces the MAPE value for the historical university data set, significantly.

Table 7: A comparison between the existing models in the enrollments dataset and the proposed model

Year	actual data	[3]	[32]	MEPA [33]	TFA [31]	[34]	[35]	Our model
1971	13055							
1972	13563	14000	14025	15430	14230	14195	14242.0	1358.4
1973	13867	14000	14568	15430	14230	14424	14242.0	13846
---	---	---	---	---	---	---	---	---
1991	19337	19000	19454	19333	18872	19000	19144.0	19891.2
1992	18876	19000	19454	19333	18872	19000	19144.0	18692
MSE		**407507**	**255227.4**	**446761.76**	**261162.3**	**261463.62**	**228920**	**57294.53**

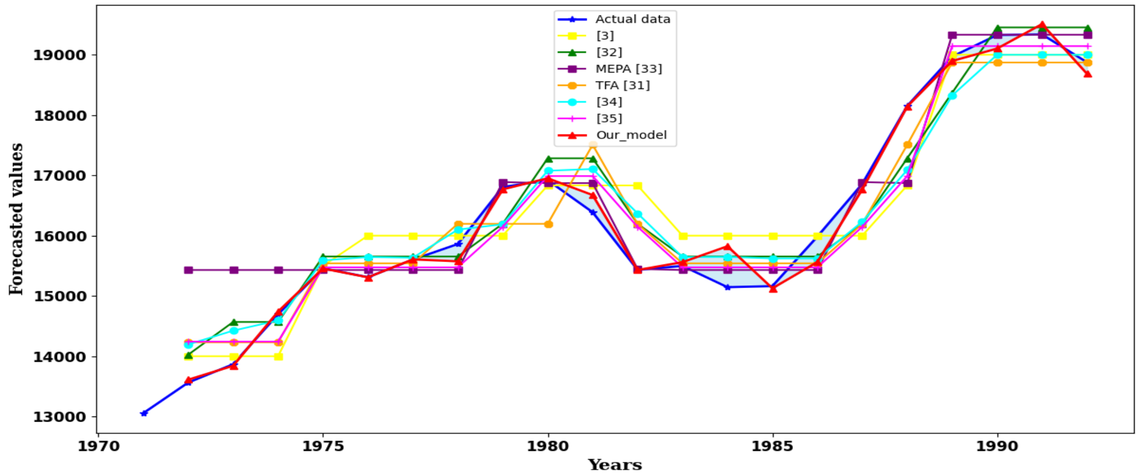

Fig. 4: The plots displaying actual data alongside those of other models and our model

In addition, the forecasted results of the proposed model are also compared with each model which is named as in works [6, 18, 4, 20] based on the various high - order FRs with different number of intervals. Comparison of these models according to the MSE (8) value is shown in Figure 5, where the CC06b model in work [20] model and the C02 model in work [4] use 7th-order and 5th- order FRs, respectively. Remaining models use fuzzy relations with number of orders less equal to 4. The results in Figure 5 confirm that our proposed model has the smallest error in comparison with four other models in terms of MSE. For the proposed model, the MSE value is 537.4 which is the smallest forecasting error as known. It can be seen that the proposed model forecasts more accurate than the existing models for various high-order models under different number of intervals

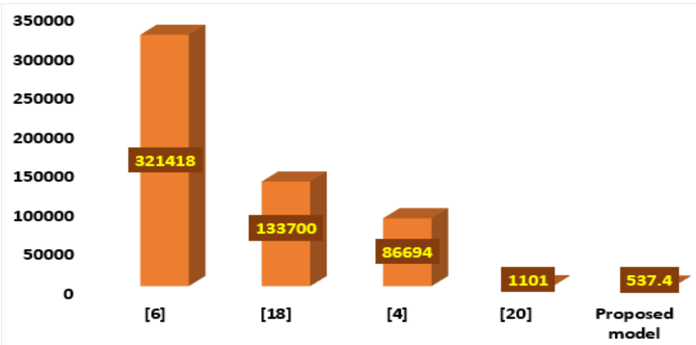

Figure 5: A comparison of the MSE value between the proposed model and various high-order FTS models

4.3. Forecasting the Gasonline price of RON 95

In this section, the proposed model is applied to forecast the Gas prices RON95 in Vietnam between 02/10/2023 and 28/03/2024. The performance of the proposed model is evaluated by using the MAPE (9). The results and accuracy of the proposed model based on different number of orders with 7 intervals which are shown in Table 8. Furthermore, the trend in forecast of the proposed method is also illustrated in Figure 6 and it clearly shows that the proposed forecasted values are significantly in close accordance with the actual values.

Table 8. The forecasting results and accuracy of the proposed model

Day	Actual data of RON95	Forecasted values
02/10/2023	24840	Not forecasted
11/10/2023	23040	22936.2
23/10/2023	23510	23532.5
----	----	----
21/03/2024	24280	24354.2
28/03/2024	24810	24564.3
MAPE		**1.02%**

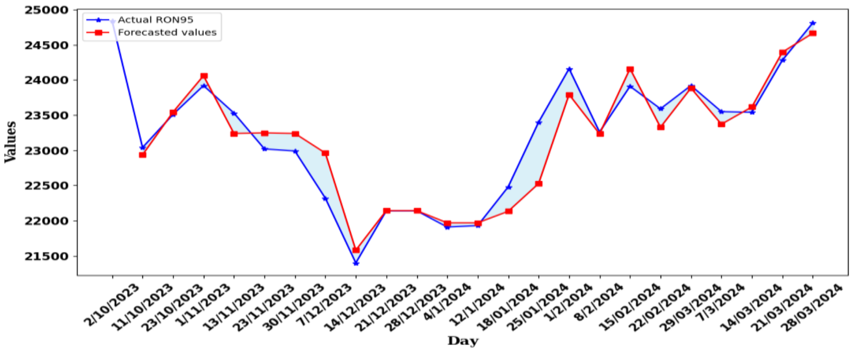

Figure 6: Depict the graph comparing the actual RON95 prices with the forecasted values of the proposed model**.**

V. CONCLUSIONS

In this paper, a FTS forecasting model utilizing a graph-based clustering technique is presented with the objective of achieving enhanced forecasting accuracy. The proposed model addresses a limitation of existing fuzzy time series forecasting approaches by automating the determination of interval lengths in the universal of discourse. Specifically, our approach employs the graph-based clustering technique to dynamically determine interval lengths, eliminating the need for pre-selecting the number of intervals by the user. Furthermore, we define time variant fuzzy relationship groups and compute forecasted values using simple computation rather than relying on the max-min composition operator on fuzzy sets. The effectiveness of proposed method is evaluated by using datasets concerning student enrollments at the University of Alabama and Gas prices RON95 in Vietnam. Our simulation and application results suggest that our approach contributes meaningfully to the literature on time series clustering.

REFERENCES

[1] Song Q, Chissom B S. Forecasting enrolments with fuzzy time series - Part I. Fuzzy Sets and Systems. 1993a;54(1):1-9.
[2] Song Q, Chissom B S. fuzzy time series and its models. Fuzzy Sets and Systems. 1993b;54(3):269-277.
[3] Chen S M. Forecasting Enrolments based on fuzzy time series. Fuzzy set and systems. 1996; 81:311-319.
[4] Chen S M. Forecasting enrolments based on high-order fuzzy time series. Cyber-netics and Systems. 2002;33(1):1-16.
[5] Huarng K. Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems. 2001;123(3):387-394.
[6] Hwang J R, Chen S M, Lee C H. Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, 1998; 100: 217–228.
[7] Yu H K. A refined fuzzy time-series model for forecasting, Physical A: Statistical Mechanics and Its Applications. 2005;346(3-4):657-681.
[8] Yu H K. Weighted fuzzy time series models for TAIEX forecasting. Phys A, Stat Mech Appl. 2005;349(3–4):609-624.
[9] Bosel M, Mali K. A novel data partitioning and rule selection technique for modelling high-order fuzzy time series. Applied soft computing. 2017, 10.1016/j.asoc.2017.11.01.
[10] Chen S M, Tanuwijaya K. Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques. Expert Systems with Applications. 2011;38: 15425–15437.
[11] Loc V M, Nghia P T H. Context-aware approach to improve result of forecasting enrollment in fuzzy time series. International Journal of Emerging Technologies in Engineering Research (IJETER). 2017;5(7):28-33.
[12] Lu. W, XueyanChen., Xiao-dongLiua. W, JianhuaYang. Using interval information granules to improve forecasting in fuzzy time series. International Journal of Approximate Reasoning. 2015; 57:1-18.
[13] Kuo I-H, et al. An improved method for forecasting enrolments based on fuzzy time series and particle swarm optimization. Expert systems with applications. 2009; 36:6108 – 6117.
[14] Kuo I-H, et al. Forecasting TAIFEX based on fuzzy time series and particle swarm optimization. Expert Systems with Applica-tions. 2010; 237:1494–1502.
[15] Tian Z H, Wang P, He T Y. Fuzzy time series based on K-means and particle swarm optimization algorithm. Man-Machine-Environement System Engineering. 2017; 406:181-189.
[16] Pattanayak, R.M.; Behera, H.; Panigrahi, S. A novel probabilistic intuitionistic fuzzy set based model for high order fuzzy time series forecasting. Eng. Appl. Artif. Intell. 99, 104136, 2021.
[17] Singh P, Borah B. An effective neural network and fuzzy time series based hybridized model to handle forecasting problems of two -factors. Knowl Inf Syst. (2014) 38: 669–690.
[18] Singh S R. A simple method of forecasting based on fuzzy time series. Applied Mathematics and Computation. 2007; 186:330–339.
[19] Huang Y L, et al. A hybrid forecasting model for enrolments based on aggregated fuzzy time series and particle swarm optimization. Expert Systems with Applications. 2011;7(38):8014–8023.
[20] 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. 2006;14: 468–477.
[21] Lee L-W, Wang, L.-H., Chen, S.-M. Temperature prediction and TAIFEX forecasting based on high-order fuzzy logical relationip and genetic simulated annealing techniques. Expert Systems with Applications. 2008;34: 328–336.
[22] Chen S M, Chung, N.Y. Forecasting enrolments using high-order fuzzy time series and genetic algorithms. International Journal of Intelligent Systems. 2006b;21: 485–501.
[23] Nai-Yi Wang, Shyi-Ming Chen. 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.
[24] S.-M. Chen, Pei-Yuan Kao, TAIEX forecasting based on fuzzy time series, particle swarm optimization techniques and support vector machines, Inf. Sci. 247 (2013) 62–71

[25] NV Tinh. Enhanced forecasting accuracy of fuzzy time series model based on combined fuzzy Cmean clustering with particle swam optimization. Int. J. Comput. Intell. Appl. 2020; 19, 2050017.

[26] Hsu L-Y, et al. Temperature prediction and TAIFEX forecasting based on fuzzy relationships and MTPSO techniques. Expert Systems with Applications. 2010;37: 2756–2770.

[27] Park J I, Lee D.J., Song C.K., Chun M.G. TAIFEX and KOSPI 200 forecasting based on two-factors high-order FTS and particle swarm optimization. Expert Systems with Applications. 2010;37: 959–967.

[28] Cheng S H, Chen S-M, Jian W S. Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures. Information Sciences. 2016;327: 272–287.

[29] Jaromczyk, J.W., Toussaint, G.T.: Relative neighborhood graphs and their relatives. Proc. IEEE 80(9), 1502–1517, 1992.

[30] Anand, R., Reddy, C.K.: Graph-based clustering with constraints. PAKDD 2011, Part II, LNAI 6635, 51–62, 2011.

[31] N.V. Tinh, T.T. Thanh, B.T.Thi, “An Interval Partitioning Approach Using Graph–Based Clustering in Fuzzy Time Series Forecasting Model,” International Journal of Multidisciplinary and Current Research, vol. 9, pp. 142-150, 2021.

[32] 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.

[33] 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.

[34] 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.

[35] 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.