East African Scholars Journal of Engineering and Computer Sciences
Abbreviated Key Title: East African Scholars J Eng Comput Sci
ISSN 2617-4480 (Print) | ISSN 2663-0346 (Online) |
Published By East African Scholars Publisher, Kenya

OPEN ACCESS

Volume-2 | Issue-5 | May-2019 |

Review Article

Forecasting Gasonline Price in Vietnam Based on Fuzzy Time Series and Automatic Clustering

Nghiem Van Tinh[1] and Nguyen Tien Duy[1]
[1]Thai Nguyen University of Technology – Thai Nguyen University, ThaiNguyen, Vietnam

*Corresponding Author
Nguyen Tien Duy

Abstract: Partitioning the universe of discourse and determining effective intervals are critical for forecasting in fuzzy time series. Equal length intervals used in most existing literatures are convenient but subjective to partition the universe of discourse. In this paper, a new forecasting model based on two computational methods, the fuzzy logical relationship groups and automatic clustering technique is presented for forecasting Gasonline Price. Firstly, we use the automatic clustering algorithm to divide the historical data into clusters and adjust them into intervals with different length. Then, based on the new obtained intervals, we fuzzify all the historical data into fuzzy sets and calculate the forecasted output by the proposed method. To illustrate the forecasting process, numerical data sets of Gasonline Prices in Viet Nam is utilized to calculate by step. The results show that the proposed model gets forecasting accuracy for a higher high - order compare with first - order fuzzy time series model.
Keywords: Forecasting, FTS, fuzzy logical relationships, automatic clustering, Gasonline prices.

INTRODUCTION

Recently, fuzzy time series has been widely used for forecasting data of dynamic and non-linear data in nature. Many previous models have been discussed for forecasting used fuzzy time series, such as enrolment (Song, Q., & Chissom, B. S. 1993; Song, Q., & Chissom, B. S. 1994; Chen, S.M. 1996; Chen, S.-M., & Chung, N.-Y. 2006a; Chen, S.M., & Chung, N.Y. 2006b), crop forecast (Singh, S. R. 2007a; Singh, S. R. 2007b), the temperature prediction (Lee, L.-W. *et al.*, 2007), stock markets (Jilani, T.A., & Burney, S.M.A. 2008), etc. There is the matter of fact that the traditional forecasting methods cannot deal with the forecasting problems in which the historical data are represented by linguistic values. (Song, Q., & Chissom, B. S. 1993; Song, Q., & Chissom, B. S. 1994) proposed the time-invariant fuzzy time and the time-variant time series model which use the max-min operations to forecast the enrolments of the University of Alabama. However, the main drawback of these methods is huge computation burden. Then (Chen, S.M. 1996), proposed the first-order fuzzy time series model by introducing a more efficient arithmetic method. After that, fuzzy time series has been widely studied to improve the accuracy of forecasting in many applications (Hwang, J. R. *et al.*, 1998). Considered the trend of the enrolment in the past years and presented another forecasting model based on the first-order fuzzy time series. At the same time (Chen, S. M. 2002; Chen, S.M., & Chung, N.Y. 2006b) proposed several forecast models based on the high-order fuzzy time series to deal with the enrolments forecasting problem (Yu, H.K. 2005) predicted the Taiwan weighted index (Liu, H.T. 2007). Presented a new forecast model based on the trapezoidal fuzzy numbers (Huarng, K.H., & Yu, T.H.K. 2006). Showed that the different lengths of intervals may affect the accuracy of forecasting. Recently (Kuo, I. H. *et al.*, 2009a), used particle swarm optimization method to search for a suitable partition of universe. Additionally (Chen, S.-M., & Tanuwijaya, K. 2011), proposed a new method to forecast enrolments based on automatic clustering techniques and fuzzy logical relationships. In this paper, we proposed a new forecasting model combining the fuzzy time series and automatic clustering technique. The method is different from the approach (Chen, S.M. 1996) and (Kuo, I. H. *et al.*, 2009a) in the way where the fuzzy relationship groups are created. Based on the model proposed in (Nghiem, V. T., & Nguyen, C. D. 2017), we have developed a new fuzzy time series model by combining the automatic clustering technique and fuzzy time series

Quick Response Code

Journal homepage:
http://www.easpublisher.com/easjecs/

Article History
Received: 05.04.2019
Accepted: 08.05.2019
Published: 28.05.2019

DOI: 10.36349/easjecs.2019.v02i05.003

with the aim to increase the accuracy of the forecasting model.

The rest of this paper is organized as follows. In Section 2, we provide a brief review of fuzzy time series and its model. In Section 3, we present a model for forecasting the Gasonline price in VietNam based on the automatic clustering technique and fuzzy time series. Then, the experimental results are shown and analyzed in Section 4. Finally, Conclusions are given in Section 5.

Preliminaries

This section introduces some important literatures, which includes fuzzy time series (Song, Q., & Chissom, B. S. 1993; Song, Q., & Chissom, B. S. 1994) and its model.

2.1. Basic Concepts of Fuzzy Time Series

Conventional time series refer to real values, but fuzzy time series are structured by fuzzy sets [1, 2]. Let $U = \{u_1, u_2, \ldots, u_n\}$ be an universal set; a fuzzy set Ai of U is defined as $A_i = \{ f_A(u_1)/u_1 +, f_A(u_2)/u_2 \ldots + f_A(u_n)/u_n \}$, where fA is a membership function of a given set A, f_A :U→[0,1], $f_A(u_i)$ indicates the grade of membership of ui in the fuzzy set A, $f_A(u_i) \in [0, 1]$, and $1 \le i \le n$.

General definitions of FTS 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 $f_i(t)$ $(i = 1,2 \ldots)$ are defined and if F(t) is a collection of $f_1(t), f_2(t), \ldots$, then F(t) is called a fuzzy time series on Y(t) (t . . ., 0, 1, 2 . ..). Here, F(t) is viewed as a linguistic variable and $f_i(t)$ represents possible linguistic values of$F(t)$.

Definition 2: Fuzzy logic relationship (FLR)

If F (t) is caused by F(t-1) only, the relationship between F(t) and F(t-1) can be expressed by $F(t-1) \rightarrow F(t)$. According to (Chen, S.M. 1996) suggested that when the maximum degree of membership of F(t) belongs to A_i, F(t) is considered A_j. Hence, the relationship between F (t) and F (t -1) is denoted by fuzzy logical relationship $A_i \rightarrow A_j$ where Ai and Aj refer to the current state or the left - hand side and the next state or the right-hand side of fuzzy time series.

Definition 3: γ - order fuzzy logical relationships

Let $F(t)$ be a fuzzy time series. If $F(t)$ is caused by $F(t-1)$, $F(t-2), \ldots, F(t-\gamma+1)$ $F(t-m)$ then this fuzzy relationship is represented by by $F(t-\gamma), \ldots, F(t-2)$, $F(t-1) \rightarrow F(t)$ and is called an m - order fuzzy time series.

Definition 4: Fuzzy relationship group (FRG)

Fuzzy logical relationships, which have the same left-hand sides, can be grouped together into fuzzy logical relationship groups. Suppose there are relationships such as follows:

$$A_i \rightarrow A_{k1}, \quad A_i \rightarrow A_{k2}, \quad \ldots\ldots$$

In previous study was proposed by Chen (1996), the repeated fuzzy relations were simply ignored when fuzzy relationships were established. So, these fuzzy logical relationship can be grouped into the same FRG as: $A_i \rightarrow A_{k1}, A_{k2} \ldots$

2.2. Fuzzy time series forecasting model

Fuzzy theory (Zadeh, L. A. 1965) was developed to deal with problems involving linguistic terms. Song and Chissom (1993; 1994) the application of this theory to define fuzzy time-series model. They applied the model to forecast the enrolments of the University of Alabama. Further, a first-order time-invariant fuzzy time series model was proposed by Song and Chissom to forecast the enrolments and a step-by-step procedure is presented to develop and utilize the time-variant model (Song, Q., & Chissom, B. S. 1994).

The evaluated fuzzy time-series model which applied simplified arithmetic operations in forecasting algorithms rather than the complicated max-min composition operations was proposed by Chen (1996). In addition, Chen's method can generate more precise forecasting results than those of Song and Chissom (1993) Chen's method consists of the following major steps:

- Step 1: Define the universe of discourse U.
- Step 2: Divide U into several equal-length intervals.
- Step 3: Define the fuzzy sets on U and fuzzify the historical data.
- Step 4: Create the fuzzy logical relationships based on the historical data.
- Step 5: Classify the derived fuzzy logical relationships into groups.
- Step 6: Forecast and defuzzify output value

A proposed forecasting model based the FTS and automatic clustering

In this section, an improved hybrid model for forecasting the Gasonline Price 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, based on the new intervals defined, we fuzzify on the historical data, determine fuzzy logical relationships and create fuzzy logical relationship groups and calculate forecasted value according to the forecasting model of Chen (1993). To verify the effectiveness of the proposed model, all historical data sets of Gasonline

Price of E5 RON 92 of Viet Nam are shown in Fig. 1, which are used to illustrate for forecasting process. The step-wise procedure of the proposed model is detailed as follows:

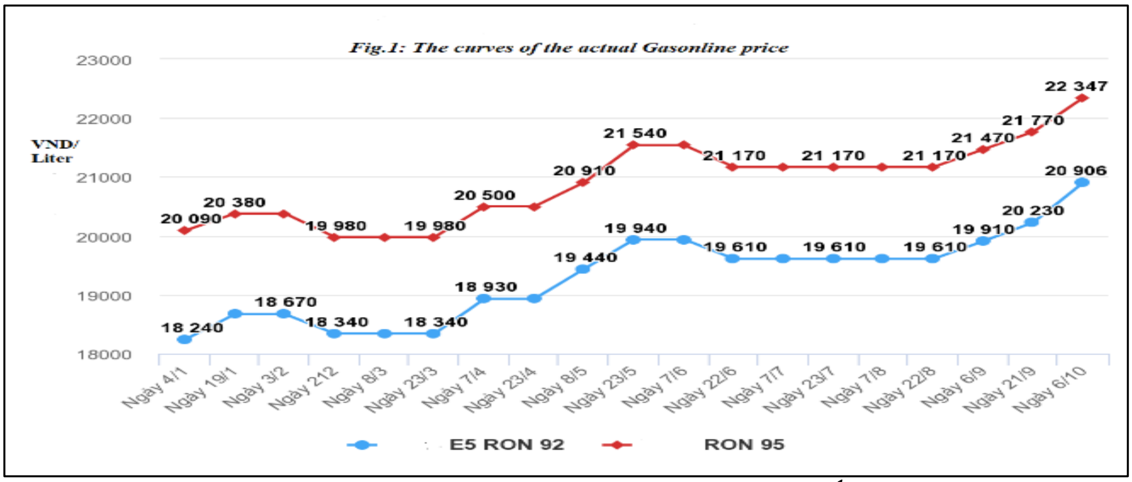

Figure.1 The curves of the actual Gasonline price[1]

Step 1: Sort the data of Gasonline prices in an ascending order. Assume that the ascending sequence of the data is shown as follows: $d_1, d_2, d_3, \ldots, d_i, \ldots, d_n$. with $d_{i-1} < d_i$; where d_1 is the smallest datum among the n numerical data d_n is the largest datum among the n numerical data, and $1 \le i \le n$.

Step 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:

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

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

Where; D denotes the mean of the data $d_1, d_2, d_3, \ldots, d_i, \ldots, d_n$.

$$D = \frac{\sum_{i=1}^{n} d_i}{n} \qquad (3)$$

Step 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_data* between any two adjacent data using $\boldsymbol{Dev_{diff}}$, shown as follows:

$$\text{Max\_data} = 0.5 * \boldsymbol{Dev_{diff}} \qquad (4)$$

Based on the value of *Max_data*, 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 di cluster as follows:

$\ldots, \{\ldots, d_i\}, d_{i+1}, d_{i+2}, \ldots, d_n,$

where $(1 \le i \le n)$

if *(* $d_{i+1} - d_i$ *) <* *Max_data)* ***then*** *put* d_i *into the current cluster in which* d_i *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 *(*$d_{i+1} - d_i$ *< 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*.

$$\text{Max\_data\_Distance} = 0.5 * \boldsymbol{Aver_{diff}} \qquad (5)$$

else *Subtract the minimum value difference min* ***(*** $d_{i+1} - d_i$ *)) between any data in the data set from the maximum value different max ((* $d_{i+1} - d_i$ *)) between any data in the data set and divide by two, use the result to calculate maximum data distance*

[1] *Source: https://kinhdoanh.vnexpress.net/tin-tuc/hang-hoa*

$$Max\_data = 0.5*(max(d_{i+1} - d_i) - min(d_{i+1} - d_i)) \quad (6)$$

Repeatedly perform the above process, until all the data have been clustered. From this viewpoint and based on steps 1, 2 and 3 aboved, we get nine clusters as follows:

C1= {18204}; C2= {18340, 18340, 18340}; C3 = {18670, 18670}; C4={18930, 18930}; C5 ={19440}; C6={19610, 19610, 19610}; C7 = {19940, 19940}; C8 ={20230}; C9 ={20906}.

Step 4: Adjust the clusters into intervals according to the following rules:
Assume that there are two adjacent clusters, clusteri and clusterk shown as follows:

$\ldots, \{d_{i1}, \ldots, d_{in}\}, \{d_{k1}, \ldots, d_{km}\}, \ldots$; Where d_{in} is the last datum in clusteri and d_{k1} is the first datum in clusterk

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

$$Cluster\_uB_i = \frac{d_{in} + d_{k1}}{2} \quad (7)$$
$$Cluster\_lB_k = Cluster\_uB_i \quad (8)$$

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:

$$Cluster\_lB_1 = d_1 - \text{Max}\_data \quad (9)$$
$$Cluster\_uB_n = d_n + \text{Max}\_data \quad (10)$$

Where d_1 the first datum is 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 $interval_k$ according to following equation:

$$Mid\_value_i = \frac{interval\_lB_i + interval\_uB_i}{2} \quad (11)$$

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.

Based on Step 4, we obtain 9 intervals corresponding to 9 clusters are shown in Table 1 as follows:

Table 1: The intervals get from automatic clustering

No	Intervals
1	u_1 = (18119.25, 18272)
2	u_2 = (18272 , 18505)
3	u_3 = (18505 , 18800)
4	u_4 = (18800 19185)
5	u_5 = (19185 19525)
6	u_6 = (19525 19775)
7	u_7 = (19775 20085)
8	u_8 = (20085 20568)
9	u_9 = (20568 20993.75)

***Step 5*: Define the fuzzy sets for observation of Gasonline Prices**

Each interval obtained in Step 4 represents a linguistic variable of "Gasonline prices". For nine intervals, there are nine linguistic values which are A_1 = "very low Gasonline prices", A_2 ="low Gasonline prices", A_3 =" above low Gasonline prices", A_4 = "average Gasonline prices", A_5 = "above average Gasonline prices", A_6 = "more high Gasonline prices", and A_7 =" high Gasonline prices", A_8 =" very high Gasonline prices and A_9=" very very high Gasonline prices to represent different regions in the universe of discourse on U, respectively. Each linguistic variable represents a fuzzy set A_i and its definitions is described in (12) as follows:

$$\begin{aligned} A_1 &= \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \cdots + \frac{0}{u_9} \\ A_2 &= \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \cdots + \frac{0}{u_9} \\ &\text{------------------------------------} \\ A_8 &= \frac{0}{u_1} + \frac{0}{u_2} + \cdots + \frac{0.5}{u_7} + \frac{1}{u_8} + \frac{0.5}{u_9} \\ A_9 &= \frac{0}{u_1} + \frac{0}{u_2} + \cdots + \frac{0}{u_7} + \frac{0.5}{u_8} + \frac{1}{u_9} \end{aligned} \qquad (12)$$

For simplicity, the membership values of fuzzy set A_i either are 0, 0.5 or 1. The value 0, 0.5 and 1 indicate the grade of membership of u_j ($1 \le j \le 9$) in the fuzzy set A_i ($1 \le i \le 9$).

Where, the symbol '+' denotes fuzzy set union, the symbol '/' denotes the membership of u_j which belongs to A_i.

Step 6: Fuzzify all historical data of Gasonline Prices

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 Gasonline data on date 21/02/2018 is 18340, and it belongs to interval u_2 because 18340 is within (18270, 18505]. So, we then assign the linguistic value ''low Gasonline prices" (eg. the fuzzy set A_1) corresponding to interval u_2 to it. Consider two time serials data Y(t) and F(t) at year t, where Y(t) is actual data and F(t) is the fuzzy set of Y(t). According to formula (11), the fuzzy set A_2 has the maximum membership value at the interval u_2. Therefore, the historical data time series on date Y (21/02/2018) is fuzzified to A_2. The completed fuzzified results of Gasonline prices are listed in Table 2.

Table 2: The complete result of fuzzification data of gasonline price

Date	Actual data	linguistic value
4/1/2018	18240	A1
19/1/2018	18670	A3
3/2/2018	18670	A3
21/2/2018	18340	A2
8/3/2018	18340	A2
23/3/2018	18340	A2
7/4/2018	18930	A4
23/4/2018	18930	A4
8/5/2018	19440	A5
23/5/2018	19940	A7
7/6/2018	19940	A7
22/6/2018	19610	A6
7/7/2018	19610	A6
23/7/2018	19610	A6
28/7/1998	19610	A6
7/8/2018	19610	A6
22/8/2018	19610	A6
6/9/2018	19910	A7
21/9/2018	20230	A8
06/10/2018	20906	A9

Step 7: Define all γ – order fuzzy logical relationships.

Based on Definition 2. To establish a γ - order fuzzy relationship, we should find out any relationship which has the $F(t-\gamma), F(t-\gamma+1), \ldots, F(t-1) \rightarrow F(t)$, where $F(t-\gamma), F(t-\gamma+1), \ldots, F(t-1)$ and $F(t)$ are called the current state and the next state of fuzzy logical relationship, respectively. Then a γ - order fuzzy logic relationship in the training phase is got by replacing the corresponding linguistic values.

For example, supposed $\boldsymbol{\gamma = 1}$ from Table 2 a first -order fuzzy relation $A_1 \rightarrow A_3$ is got as $F(4/12018) \rightarrow F(19/1/2018)$. So on, we get the first-order fuzzy relationships are shown in Table 3.

Table 3: The complete result of the first- order fuzzy logical relationship groups

No	linguistic value	First – order fuzzy relations
1	A1	A1 → A3
2	A3	A3 → A3
3	A3	A3 → A2
4	A2	A2 → A2
5	A2	A2 → A2
--	--	--------
17	A6	A6 → A6
18	A7	A6 → A7
19	A8	A7 → A8
20	A9	A8→ A9

Step 8: Establish all γ – order fuzzy logical relationships groups

Based on (Nghiem, V. T., & Nguyen, C. D. 2017) all the fuzzy relationships having the same fuzzy set on the left-hand side or the same current state can be put together into one fuzzy relationship group. The repeated fuzzy logical relationships are also considered up to the forecasting time. Therefore, from Table 3 , we can obtain 20 first – order fuzzy relationship groups shown in Table 4.

Table 4: The complete result of the first- order fuzzy logical relationship groups

No	**linguistic value**	**First – order fuzzy relation groups**
1	A1	A1 → A3
2	A3	A3 → A3
3	A3	A3 → A3,A2
4	A2	A2 → A2
--	--	-------------
16	A6	A6 → A6, A6, A6,A6
17	A6	A6 → A6, A6, A6, A6,A6
18	A7	A6 → A6, A6, A6, A6,A6, A7
19	A8	A7 → A7, A6, A8
20	A9	A8→ A9

Step 9: Calculate and defuzzify the forecasted output values

In this step, to obtain the forecasted results, we use defuzzification technique [20] to calculate the forecasted values for all date of Gasonline price in training phase. The defuzzification rules are presented for date t (04/01/2018 ≤ t ≤ 06/10/2018) as follows:

Principle 1: If the fuzzified data of Gasonline price on date t-1 is A_j and there is only one fuzzy logical relationship in the fuzzy logical relationship group whose current state isA_j, shown as follows: $A_j(t-1) \rightarrow A_k(t)$, then the forecasted Gasonline price of datet is m_k, where mk is the midpoint of the interval uk and the maximum membership value of the fuzzy set A_k occurs at the interval u_k.

Principle 2: If the fuzzified Gasonline price of datet t -1 is A_j and
There is the following fuzzy logical relationship group whose current state is Aj, shown as follows:

$$A_j(t-1) \rightarrow A_{i1}(t1), A_{i2}(t2), A_{ip}(tk)$$

Then the forecasted Gasonline price of datet t is calculated as follows:

$$\text{forecasted} = \frac{1*m_{i1}+2*m_{i2}+3*m_{i3}+\cdots+p*m_{ip}}{1+2+\cdots+p};$$

Where m_{i1}, m_{i2}, m_{ik} is the middle values of the intervals ui1, ui2 and uip respectively, and the maximum membership values of $A_{i1}, A_{i2}, \dots, A_{ip}$ occur at intervals$u_{i1}, u_{i2}, \dots, u_{ip}$, respectively.

Based on the forecasted rules above and based on Table 4, we complete forecasted results of Gasonline price in Viet Nam based on first - order FTS model with nine intervals are listed in Table 5.

Table 5: The complete forecasted outputs for gasonline price of Viet Nam based on the first– order FTS model

Date	**Actual data**	**linguistic value**	**Forecasted value**
4/1/2018	18240	A1	----------
19/1/2018	18670	A3	18652.5
3/2/2018	18670	A3	18652.5
21/2/2018	18340	A2	18476.5
8/3/2018	18340	A2	18388.5
23/3/2018	18340	A2	18388.5
7/4/2018	18930	A4	18690.5
23/4/2018	18930	A4	18992.5
8/5/2018	19440	A5	19234.17
23/5/2018	19940	A7	19930
7/6/2018	19940	A7	19930
22/6/2018	19610	A6	19743.33
7/7/2018	19610	A6	19650
23/7/2018	19610	A6	19650
28/7/1998	19610	A6	19650
7/8/2018	19610	A6	19650
22/8/2018	19610	A6	19650
6/9/2018	19910	A7	19730
21/9/2018	20230	A8	20034.92
06/10/2018	20906	A9	20780.88

The performance of proposed model can be assessed by comparing the difference between the forecasted values and the actual values. The mean square error (MSE) is employed as an evaluation criterion to represent the forecasted accuracy. The MSE value is computed as follows:

$$MSE = \frac{1}{n}\sum_{t=\gamma}^{n}(F_t - R_t)^2 \quad (13)$$

Where, R_t denotes actual value at year t, F_t is forecasted value at year t, n is number of the forecasted data, γ is order of the fuzzy logical relationships

Computing Results

In this paper, we apply the proposed model to forecast the gasonline price of Viet Nam with the whole historical data the period from 4/1/2018 to 6/10/2018 is shown in Fig.1. We implement the proposed model under different number of orders and kept number of intervals of 9. The forecasted accuracy of the proposed method is estimated by using the MSE function (12). The forecasted results of proposed model under number of interval as 9 and various order of fuzzy relationship are listed in Table 6.

Table 6: The completed forecasting results for gasonline price data of Viet Nam under deferent orders of FTS

Date	Actual data	The forecasted value for each order of fuzzy logical relationship					
		2nd-order	*3rd-order*	*4th-order*	*5th-order*	*6th-order*	*7th-order*
4/1/2018	18240	-----	-----	-----	-----	-----	-----
19/1/2018	18670	-----	-----	------	------	-----	------
3/2/2018	18670	18652.5	-----	-----	------	-----	------
21/2/2018	18340	18388.5	18388.5	----	------	-----	------
8/3/2018	18340	18388.5	18388.5	18388.5	-----	-----	-----
23/3/2018	18340	18388.5	18388.5	18388.5	18388.5	-----	-----
7/4/2018	18930	18791.17	18992.5	18992.5	18992.5	18992.5	-----
23/4/2018	18930	18992.5	18992.5	18992.5	18992.5	18992.5	18992.5
8/5/2018	19440	19355	19355	19355	19355	19355	19355
23/5/2018	19940	19930	19930	19930	19930	19930	19930
7/6/2018	19940	19930	19930	19930	19930	19930	19930
22/6/2018	19610	19650	19650	19650	19650	19650	19650
7/7/2018	19610	19650	19650	19650	19650	19650	19650
23/7/2018	19610	19650	19650	19650	19650	19650	19650
28/7/1998	19610	19650	19650	19650	19650	19650	19650
7/8/2018	19610	19650	19650	19650	19650	19650	19650
22/8/2018	19610	19650	19650	19650	19650	19650	19650
6/9/2018	19910	19743.33	19762	19790	19836.67	19930	19930
21/9/2018	20230	20326.5	20326.5	20326.5	20326.5	20326.5	20326.5
06/10/2018	20906	20780.88	20780.88	20780.88	20780.88	20780.88	20780.88
MSE		**5573**	**4633**	**4307**	**3836**	**3536**	**3561**

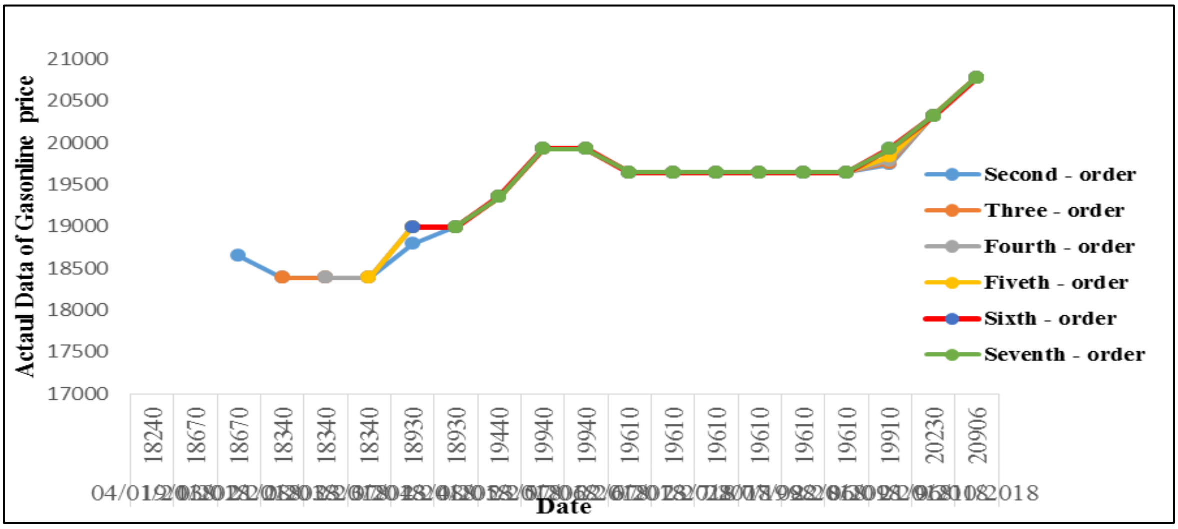

Figure. 2. A comparison of the MSE values for 9 intervals with different high-order fuzzy relationships

From Table 6, it can be seen that the performance of the proposed model is improved a lot with increasing number of orders in the same number of interval. Particularly, the proposed model gets the lowest MSE value of ***3536*** with the 6th-order fuzzy logical relationship. The trend in forecasting of Gasonline price by the high-order fuzzy time series model in comparison to the actual data with the different number of orders can be visualized in Fig.2.

CONCLUSIONS

In this paper, a new hybrid forecasting model based on combined FTS and Automatic clustering technique is presented to improve forecasting. First, the proposed model is implemented for forecasting of rice production of Viet Nam. Subsequently, we proved correctness and robustness of the proposed model by testing and verifying it on historical gasonline price data of Vietnam and comparing with different number of fuzzy relationship in the training phase. The main contributions of this paper are illustrated in the following. First, the author shows that the forecasted accuracy of the proposed model can be improved by considering more information of latest fuzzy fluctuation within all current states of all fuzzy relationships. Second, the Automatic clustering algorithm for the optimized lengths of intervals is developed to find optimal interval. Third, based on the performance comparison in Table 6 and Fig.2 the author shows the proposed model outperforms with increasing number of orders in the same number of interval. To continue considering the effectiveness of the forecasting model in the future, the proposed model can be extended to deal with multidimensional time series data such as: Stock market prediction, weather forecast, traffic accident prediction and so on.

Acknowledge

This paper was implemented by the "Knowledge technologies and soft computing – TNUT" research group.

REFERENCES

1. Song, Q., & Chissom, B. S. (1993). Forecasting enrollments with fuzzy time series—part I. *Fuzzy sets and systems*, *54*(1), 1-9.
2. Song, Q., & Chissom, B. S. (1994). Forecasting enrollments with fuzzy time series—part II. *Fuzzy sets and systems*, *62*(1), 1-8.
3. Chen, S.M. (1996). "Forecasting Enrollments based on Fuzzy Time Series", *Fuzzy set and system*, 81, 311-319.
4. Hwang, J. R., Chen, S. M., & Lee, C. H, (1998). "Handling forecasting problems using fuzzy time series", *Fuzzy Sets and Systems*, 100(1-3), 217-228.
5. Huarng, K. (2001b). "Heuristic models of fuzzy time series for forecasting" *Fuzzy Sets and Systems*, 123, 369-386.
6. Singh, S. R. (2007a). "A simple method of forecasting based on fuzzytime series", *Applied Mathematics and Computation*, 186, 330-339.
7. Singh, S. R. (2007b). "A robust method of forecasting based on fuzzy time series", *Applied Mathematics and Computation*, 188, 472-484.
8. Chen, S. M. (2002). "Forecasting enrollments based on high-order fuzzy time series", *Cybernetics and Systems: An International Journal*, 33, 1-16.
9. Yu, H.K. (2005). "Weighted fuzzy time series models for TAIEX forecasting", Physica A, 349, pp. 609-624.
10. Chen, S.-M., & Chung, N.-Y. (2006a). "Forecasting enrollments of students by using fuzzy time series and genetic algorithms", *International Journal of Information and Management Sciences* 17, 1-17.
11. Chen, S.M., & Chung, N.Y. (2006b). "Forecasting enrollments using high-order fuzzy time series and genetic algorithms" *International of Intelligent Systems* 21, 485-501.
12. Huarng, K.H., & Yu, T.H.K. (2006). "Ratio-Based Lengths of Intervals to Improve Fuzzy Time Series Forecasting," *IEEE Transactions on SMC - Part B: Cybernetics*, 36, 328-340.
13. Liu, H.T. (2007). "An Improved fuzzy Time Series Forecasting Method using Trapezoidal Fuzzy Numbers", *Fuzzy Optimization Decision Making*, 6, 63-80.
14. Lee, L.-W., Wang, L.-H., & Chen, S.-M. (2007). "Temperature prediction and TAIFEX forecasting based on fuzzy logical relationships and genetic algorithms", *Expert Systems with Applications*, 33, 539-550.
15. Jilani, T.A., & Burney, S.M.A. (2008). "A refined fuzzy time series model for stock market forecasting", Physica A 387, 2857-2862.
16. Wang, N.-Y., & Chen, S.-M. (2009). "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.
17. Kuo, I. H., Horng, S.-J., Kao, T.-W., Lin, T.-L., Lee, C.-L., & Pan, (2009a). "An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization", *Expert Systems with applications*, 36, 6108-6117.
18. Chen, S.-M., & Tanuwijaya, K. (2011). "Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques", *Expert Systems with Applications* 38, 15425-15437.
19. Zadeh, L. A. (1965). "Fuzzy sets", *Information and Control*, 8, 338-353.
20. Nghiem, V. T., & Nguyen, C. D. (2017). "Improving the forecasted accuracy of model based on fuzzy time series and k-means clustering, *Journal of Science and Technology: issue on Information and Communications Technology*, (2), 51-60.