American Journal of Information Science and Computer Engineering
Vol. 5, No. 3, 2019, pp. 103-112
http://www.aiscience.org/journal/ajisce
ISSN: 2381-7488 (Print); ISSN: 2381-7496 (Online)

A Novel Forecasting Model Combining the High – Order Fuzzy Time Series with Particle Swam Optimization

Nghiem Van Tinh*, Nguyen Tien Duy

Faculty of Electronics, Thai Nguyen University of Technology-Thai Nguyen University, Thai Nguyen, Viet Nam

Abstract

This paper proposes a novel fuzzy time series forecasting model for contributing to the stages of determining of interval lengths, establishing of fuzzy logical relationships (FLRs) and the stage of defuzzification. In fuzzy time series (FTS) models, lengths of intervals always affect the results of forecasting. Therefore, we use particle swarm optimization (PSO) technique to find the optimal length of intervals in the universe of discourse. Most of the existing forecasting models simply ignore the repeated FLRs without any proper justification or accept the number of recurrence of the FLRs without considering the appearance history of these fuzzy sets in the grouping fuzzy logical relationships process. Therefore, in this study, we consider the appearance history of the fuzzy sets on the right-hand side of the FLRs to establish the high – order fuzzy logical relationship groups, called the high-order time-variant fuzzy relationship groups (TV-FRGs) and then, a new forecasting computational technique in the stage of defuzzification is introduced with the intend to obtain the smallest forecasting error as possible. For verifying the suitability of proposed method, two numerical datasets about enrollments of the University of Alabama and Gasonline Price in Viet Nam are illustrated for forecasting process and comparing with other forecasting model. Experimental results show that the proposed model outperforms other baseline forecasting model based on the high-order FTS.

Keywords

Forecasting, Fuzzy Time Series, Fuzzy Logical Relationships, PSO, Enrollments

Received: August 25, 2019 / Accepted: October 31, 2019 / Published online: November 21, 2019

1. Introduction

With an ever increasing population and the need to provide quality service delivery, comes the necessity to budget and make estimates. From past data, projections can be made about the future and plans made to provide the necessary services that would enhance outputs. Forecasting is prediction of future events based on past or present experiences. It plays important role in our daily life as it is applied in forecasting a lot of events including weather prediction, gas price forecasting, enrollments forecasting, stock index, population growth, etc. To make a forecast with 100% accuracy for these events may not be possible, but, one can improved their accuracy and forecast processing speed. In order to solve forecasting problems in which the historical data are represented by linguistic values, many researchers proposed several different models based on fuzzy time series concept. In this study, we also present a novel refined forecasting model, which is developed by combining fuzzy time series theory with PSO. The major objective of designing such a hybridized model is explained as follows:

For fuzzification of time series data set, determination of lengths of intervals of the historical time series data set is very important. In most of the fuzzy time series models [2, 11, 12, 22-24, 27], the lengths of the intervals were kept the same. No precise reason is mentioned for using the fix lengths of

* Corresponding author
E-mail address: nghiemvantinh@tnut.edu.vn (N. V. Tinh)

intervals. K. Huarng [4] pointed out that the different lengths of intervals in the universe of discourse can affect the forecasting result and a proper choice of the length of each interval can greatly improve the forecasting accuracy rate. So, for creating the effective lengths of intervals, the particle swarm optimization technique is used to adjust the initial length of each interval in the universe of discourse by minimizing MSE value in this paper.

After generating the intervals, time series data set is fuzzified based on the fuzzy time series theory. Each fuzzified historical data values are then used to create the FLRs. Still most of the existing fuzzy time series models ignore repeated FLRs [2]. To explain this, consider the following example, suppose there are FLRs at three different time functions as follows:

$$F(t = 1)\ A_i \rightarrow A_j$$
$$F(t = 2)\ A_i \rightarrow A_k$$
$$F(t = 3)\ A_i \rightarrow A_j$$

In above FLRs, there two FLRs at time t=1 and t=3 have the same fuzzy set on the right-hand side (A_j) According to [2], these FLRs can be represented in the following FRG as $A_i \rightarrow A_j, A_k$. These fuzzy time series models do not consider the identical FLRs during forecasting. They simply use the FRGs by discarding the repeated FLRs. The ignorance of repeated FLRs in the FRG is not properly justified. Another approach, Yu [28] argued that recurrent fuzzy logical relationships should be considered during forecasts. From example above, the FLR $A_i \rightarrow A_j$ occurs two times, but it can still be grouped into the same FLRG as: $A_i \rightarrow A_j, A_k, A_j$. Yu said that the recent FLRs as more important than the previously ones. However, this scheme of establishing fuzzy logical relationship is not justifiable for each of forecasting time, because it does not consider the appearance history of the fuzzy sets on the right-hand side of the FLRs. To resolve these problems, we use a method [19] for establishing fuzzy relationship group in the model. In this approach, the FRGs are determined by considering the appearance history of the fuzzy sets on the right-hand side of the FLRs. That is, only the fuzzy sets on the right-hand side appearing before the fuzzy sets on the left-hand side of the FLR at forecasting time is put together to form FRG. In order to explain this, reviewing the fuzzy relationships mentioned above.

Suppose the forecasting time is t of 2, fuzzy sets on the right-hand side of FLRs with the same current state (right-hand side of FLRs) is considered to put into the same as G1 as follows: $A_i \rightarrow A_j, A_k$, that is, only FLRs appearing before time t is grouped into a same group. The same way, if at the time t=3, the FLRs having the same current state are grouped into a same fuzzy logical relationship group G2 as follow: $A_i \rightarrow A_j, A_k, A_j$. The advantage of using such approach is that the model can capture more persuasive information of the FLRs based on their chronological order. Although this study shows that the superior forecasting capability compared with previous forecasting models, we still continue improving the effectiveness of the forecasting model by introducing a new defuzzification technique in this paper. Furthermore, many researchers show that high-order FLRs improve the forecasting accuracy of the models [3, 5 15-18]. Therefore, in this study, we employ the high-order fuzzy relationships for obtaining the forecasting results. The proposed model has the advantage that it can produce good forecasting results. We verify the effectiveness of the proposed model using the following two real-world data set: (1) university enrollments dataset of Alabama [23], (2) Historical data of the TAIFEX [17] in Taipei, Taiwan. The empirical study on the enrolment data at the University of Alabama and the stock market dataset of TAIFEX show that the performance of proposed model is better than those of any existing models.

The rest of this paper is organized, as follows: Section 2 presents related works for FTS models. In Section 3, a brief review of the basic concepts of FTS and algorithms are introduced. In Section 4, an improved forecasting model based on the high – order TV-FRGs and PSO algorithm is demonstrated. Section 5 compares and evaluates the forecasting performance of the proposed model with the existing methods on the enrolment data of the University of Alabama and the Gasonline price data. Finally, conclusion and future work are discussed in Section 6.

2. Related Works

In the past decades, many forecasting models have were proposed to deal with various prediction problems such as: forecasting university enrolments, temperature prediction, crop productions, sales forecasting, car road accidents, inventory demand forecasting, etc. Based on the fuzzy time series theory, first forecasting model was introduced by Song and Chissom [23, 24], which were used to forecast the time series values based on linguistic values. They presented the fuzzy time series model by means of fuzzy relational equations involving max–min composition operation, and applied the model for forecasting the enrollments in the University of Alabama. Unfortunately, their method had many drawbacks such as huge computation when the fuzzy rule matrix is large and lack of persuasiveness in determining the universe of discourse and the length of intervals. Therefore, Chen [2] proposed the first-order fuzzy time series model by using simple arithmetic calculations instead of max-min composition operations [23] for better forecasting accuracy. Later, many studies provided some improvements in Chen [2] method in terms of following issues: determining of lengths of intervals, fuzzification, fuzzy logical relationships, defuzzification techniques.

To further enhance forecasting accuracy of model, many researchers proposed various fuzzy time series models. For example, Huarng [3] presented an effective approach which can properly adjust the lengths of intervals. He pointed out that the different lengths of intervals in the universe of discourse can affect the forecasting result and a proper choice of the length of each interval can greatly improve the forecasting accuracy rate. Researchers [17, 27] proposed computational methods of forecasting based on the high-order FLRs to overcome the drawback of fuzzy first-order forecasting models [2, 11]. Singh [22] proposed a new method in fuzzy time series forecasting. This model has the advantage that it minimizes the time of complicated computations of fuzzy relational matrices or to find the steady state of fuzzy relational matrix. Huarng et al. [12] exploited neural networks to construct FTS model. Their model was used to forecast stock index and obtained better forecasting result.

In recent years, many researchers have proposed various artificial intelligence based models to solve complex problems in forecasting. For example, Lee et al. [17] presented method for forecasting the temperature and the TAIFEX based on the high – order FRGs and genetic algorithm. They also used simulated annealing techniques [16] to adjust the length of each interval in the universe of discourse for increasing the forecasting accuracy rate. By introducing genetic algorithm, Chen & Chung [5] presented the high-order FTS model to forecast the enrollments of students of University of Alabama. Particle swarm optimization technique has been successfully applied in many applications for partition the universe of discourse as can be found in [4, 7, 8, 18-21]. Based on Chen's model [2], Kuo et al. [14] introduced a new hybrid forecasting model which combined fuzzy time series with PSO algorithm to find the proper length of each interval. Then, to improve method in [14], Kuo et al. [15] presented a new hybrid forecast method to solve the TAIFEX forecasting problem based on FTS and PSO algorithm. In addition, Hsu et al. [18] proposed a new method for the temperature prediction and the TAIFEX forecasting, based on two-factor high-order fuzzy logical relationships and particle swarm optimization. Huang et al. [10] proposed the hybrid forecasting model based on PSO algorithm and the refinement which aggregates the global information of fuzzy relationships with the local information of latest fuzzy fluctuation to calculate the forecasted value in the forecasting stage. In addition, Dieu N. C et al. [20] introduced the concept of time-variant fuzzy logical relationship groups and used it in the determining of logical fuzzy relationship stage for the forecasting problems. Employing the PSO techniques to find optimal intervals in partitioning the universe of discourse, the experimental results showed a significant improvement in forecasting accuracy of the proposed model. The forecasting model in [19] is also based on PSO technique and TV-FRGs but extended on the case of high order TV-FRGs to forecast stock market indices of TAIFEX. Park et al [21]. Present a new method using PSO and two-factors high-order FTS with the aim to increase the forecasting accuracy. Chen and Bui [4] use the PSO technique not only to obtain optimal partition of intervals but also to obtain optimal weight vectors. They proposed the forecasting model based on optimal partitions of intervals in the universe of discourse and optimal weighting vectors of two-factors second-order fuzzy-trend logical relationship groups simultaneously for forecasting the TAIFEX and the NTD/USD exchange rates. Cheng et al. [8] presented a fuzzy forecasting method for forecasting the TAIFEX based on FTS, fuzzy logical relationships (FLRs), PSO techniques, the K-means clustering algorithm and similarity measures between the subscripts of fuzzy sets, where the K-means clustering algorithm is used to get the cluster centers of subscripts of fuzzy sets and basing in which to formulate FRGs for forecasting the TAIFEX. Chen and Jian [7] developed a fuzzy forecasting algorithm based on PSO and similarity measures between the subscripts of fuzzy sets. Another approach for determining optimal intervals is considered as the main factor affecting the performance of forecasting such as, clustering methods of rough set [1], fuzzy C-means [9], automatic clustering [6], K-means [25, 26].

3. Basic Concepts of FTS and Algorithms

3.1. Basic Concepts of FTS

This section briefly introduces the concepts and notations related to fuzzy time series as follows. Let $U = \{u_1, u_2, \ldots, u_n\}$ be an universal set; a fuzzy set A of U is defined as $A = \{ f_A(u_1)/u_1 +, f_A(u_2)/u_2 \ldots + f_A(u_n)/u_n \}$, where fA is the membership function of A, f_A: U→[0, 1], $f_A(u_i)$ is the degree of membership of the element ui in the fuzzy set A. Here, the symbol "+" indicates the operation of union and the symbol "/" indicates the separator rather than the commonly used summation and division in algebra. The definition of FTS is briefly reviewed as follows:

Definition 1: Fuzzy time series [23, 24]

Let Y (t) (t=.., 0, 1, 2..), a subset of R, be the universe of discourse on which fuzzy sets f_i (t) (i=1, 2…) are defined and if F (t) is a collection of f_1 (t), f_2 (t),…, 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).

Example: The common observations of the performance of a student during the final year of degree examination can be represented using the fuzzy sets "good", "very good", "poor", "bad", "very bad", etc. All these words can be represented by fuzzy sets.

Definition 2: Fuzzy logic relationship (FLR) [23, 24]

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)→F (t). According to [2] suggested that when the maximum degree of membership of F (t) belongs to A_i, F (t) is considered A_k. Hence, the relationship between F (t) and F (t-1) is denoted by fuzzy logical relationship $A_i \rightarrow A_k$ where Ai and A_k 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: α-order fuzzy logical relationships [23, 24]

Let F (t) be a fuzzy time series. If F (t) is caused by F (t-1), F (t-2),…, F (t-α+1), F (t-α) then this fuzzy relationship is represented by F (t-α),…, F (t-2), F (t-1)→F (t) and is called an m-order fuzzy time series.

Definition 4: Fuzzy relationship group (FRG) [1]

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: $A_i \rightarrow A_{k1}$, $A_i \rightarrow A_{k2}$,…, $A_i \rightarrow A_{km}$.

In previous study was proposed by Chen [4], 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: $A_i \rightarrow A_{k1}, A_{k2}, \ldots, A_{km}$.

3.2. The PSO Algorithm

PSO was first introduced by Eberhart and Kannedy [13] in 1995, is a random searching algorithm based on group cooperation that can solve the near optimal solution of any kind of optimization problems [14] and is inspired by simulating the social behaviour of animals, such as fish schooling, birds flocking and the swarm theory. It is particle swarm optimization initializes each particle randomly, and then finds the optimal solution through iteration. At each step of optimization, the particles update themselves by tracking their own best position and the best particle. To get the optimal solution, the particles update their own speed and positions according to the following formulas:

$$V_{id}^{k+1} = \omega^k \times V_{id}^k + C_1 \times Rand() \times (P_{best\_id} - X_{id}^k) + C_2 \times Rand() \times (G_{best} - X_{id}^k) \quad (1)$$

$$X_{id}^{k+1} = X_{id}^k + X_{id}^{k+1} \quad (2)$$

$$\omega^k = \omega_{max} - \frac{k \times (\omega_{max} - \omega_{min})}{iter\_max} \quad (3)$$

where,

1. X_{id}^k is the current position of a particle id in k-th iteration;
2. V_{id}^k is the velocity of the particle id in k-th iteration, and is limited to $[-V_{max}, V_{max}]$, where V_{max} is a constant pre-defined by user.
3. P_{best_id} is the position of the particle id that experiences the best fitness value.
4. G_{best} is the best one of all personal best positions of all particles within the swarm.
5. Rand () is the function can generate a random real number between 0 and 1 under normal distribution.
6. C_1 and C_2 are acceleration values which represent the selfcondence coefficient and the social coefficient, respectively.
7. ω is the inertia weight factor accoding to Eq. (3).

4. A Refined Forecasting Model Based on the High – Order FTS and PSO Algorithm

In this section, a novel refined forecasting model which combined the high – order TV-FRGs and PSO algorithm is introduced. The detail of the proposed model is presented as follows:

4.1. Forecasting Model Based on the High – Order FTS

To verify the effectiveness of the proposed model, all historical enrollments data [2] from 1971s to 1992s are used to illustrate the high-order FTS forecasting process. The step-wise procedure of the proposed model is detailed as follows:

Step 1: Define the universe of discourse U

Assume Y (t) be the historical data of enrolments at year t ($1971 \le t \le 1992$). The university of discourse is defined as $U = [D_{min}, D_{max}]$. In order to ensure the forecasting values bounded in the universe of discourse U, we set $D_{min} = I_{min} - N_1$ and $D_{max} = I_{max} + N_2$; where I_{min} , I_{max} are the minimum and maximum data of $Y(t)$; N_1 and N_2 two proper positive real values denote the buffers to adjust the lower bound and the upper bound of the universe of discourse and to cover the noise of the testing data.

From the historical data, we obtain $I_{min} = 13055$ và $I_{max} = 19337$. Thus, the universe of discourse is defined as U= $[I_{min} - N_1, I_{max} + N_2]$ =[13000, 20000] with $N_1 = 55$ and $N_2 = 663$.

Step 2: Partition U into equal length intervals

Divide U into equal length intervals. Compared to the previous models in article [2, 14] and for convenience to demo of the forecasting example here, we firstly divide U into seven intervals, $u_1, u_2, \dots, u_7$, respectively. The length of each interval is $L = \frac{D_{max} - D_{min}}{7} = \frac{20000-13000}{7} = 1000$. Thus, the seven intervals are defined as follows:

u_i=[13000 + (i-1)×L, 13000 + i×L), with $(1 \le i \le 7)$ gets seven intervals as:

u_1=[13000, 14000), u_2=[14000, 15000),…, u_6=[18000, 19000), u_7=[19000, 20000).

Step 3: Define the fuzzy sets for each interval

Each of interval in Step 2 represents a linguistic variable of "enrolments" in [2]. For seven intervals, there are seven linguistic values which are A_1="not many", A_2="not too many", A_3="many", A_4 ="many many", A_5 ="very many", A_6 ="too many", and A_7="too many many" 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 (4) and (5) as follows.

$$A_i = {a_{i1}}/{u_1} + {a_{i2}}/{u_2} + \dots + {a_{ij}}/{u_j} + \dots + {a_{i7}}/{u_7} \qquad (4)$$

$$a_{ij} = \begin{cases} 1 \; j = i \\ 0.5 \; j = i-1 \; or \; j = i+1 \\ 0 \; otherwise \end{cases} \qquad (5)$$

Here, the symbol '+' denotes the set union operator, $a_{ij} \in [0, 1]$ $(1 \le i \le 7,\ 1 \le j \le 7)$, u_j is the j^{th} interval of U. The value of a_{ij} indicates the grade of membership of u_j in the fuzzy set A_i. For simplicity, the different membership values of fuzzy set A_i are selected by according to Eq. (5). According to Eq. (4) and (5), a fuzzy set contains 7 intervals. Contrarily, an interval belongs to all fuzzy sets with different membership degrees. For example, u_1 belongs to A_1 and A_2 with membership degrees of 1 and 0.5 respectively, and other fuzzy sets with membership degree is 0.

Step 4: Fuzzy all historical enrolments 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 1972 is 13563, and it belongs to interval u_1 because 13563 is within [13000, 14000). So, we then assign the linguistic value "not many" (eg. the fuzzy set A_1) corresponding to interval u_1 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 Eq. (4), the fuzzy set A_1 has the maximum membership value at the interval u_1. Therefore, the historical data time series on date Y (1972) is fuzzified to A_1. The completed fuzzified results of the enrollments are listed in Table 1.

Table 1. The results of fuzzification for enrollments data.

Year	Actual data	Fuzzy sets	Linguistic value
1971	13055	A_1	"not many"
1972	13563	A_1	"not many"
1991	19337	A_7	"too many many"
1992	18876	A_6	"too many"

Step 5: Define all α – order fuzzy logical relationships.

Based on Definition 3. To establish the α-order fuzzy relationship, we should find out any relationship which has the $F(t-\alpha), F(t-\alpha+1), \dots, F(t-1) \rightarrow F(t)$, where $F(t-\alpha), F(t-\alpha+1), \dots, F(t-1)$ and $F(t)$ are called the current state and the next state of fuzzy logical relationship, respectively. Then a α-order fuzzy relationship in the training phase is got by replacing the corresponding linguistic values as follows: $A_{i\alpha}, A_{i(\alpha-1)}, \dots, A_{i2}, A_{i1} \rightarrow A_k$. For example, supposed $\alpha = 3$, a fuzzy relationship $A_1, A_1, A_1 \rightarrow A_2$ is got as $F(1971), F(1972), F(1973) \rightarrow F(1974)$. Continue with examples above and based on Table 1, all 3rd-order FLRs are shown in Table 2.

Table 2. The complete results for the first-order FLRs.

Year	No	Fuzzy relations
1974	1	$A_1, A_1, A_1 \rightarrow A_2$
1975	2	$A_1, A_1, A_2 \rightarrow A_3$
1992	19	$A_4, A_7, A_7 \rightarrow A_6$
1993	20	$A_7, A_7, A_6 \rightarrow \#$

Step 6: Establish all α – order fuzzy relationships groups

In previous studies [2, 14] all the fuzzy logical relationships having the same fuzzy set on the left-hand side or the same current state can be grouped into a same fuzzy logical relationship group. But, according to article [19], the appearance history of the fuzzy sets on the right-hand side of fuzzy logical relationships is need to more consider. That is, only the element on the right-hand side appearing before the left-hand side of the fuzzy logical relationship is put into the same fuzzy logic relationship group. For example, suppose that there two 3rd – order fuzzy logical relationships with the same left – hand side as follows: $A_i, A_j, A_k \rightarrow A_p$; $A_i, A_j, A_k \rightarrow A_q$. These fuzzy logical relationships can be grouped together into two group G1 and G2 in chronological order are listed as follows: G1: $A_i, A_j, A_k \rightarrow A_p$ or G2: $A_i, A_j, A_k \rightarrow A_p$, A_q. From this viewpoint and based on Table 2, we can obtain 20 the high-order TV-FRGs are shown in column 5 of Table 3.

Where, the first 19 groups of the high – order fuzzy logical relationships are called the trained patterns (or in training phase), and the last one is called the untrained pattern (or in testing phase).

Table 3. The complete results for the third-order FRGs.

Year	Fuzzy set	No group	Third-FRGs
1974	A_2	G1	$A_1, A_1, A_1 \rightarrow A_2$
1992	A_6	G19	$A_4, A_7, A_7 \rightarrow A_6$
1993	N/A	G20	$A_7, A_7, A_6 \rightarrow \#$

Step 7: Calculate and defuzzify the forecasted values

To defuzzify the fuzzified data and to obtain the forecasted values, a new defuzzification technique is developed to calculate the forecasted values for all high – order TV-FRGs in training phase. Then we use the master voting (MV) scheme for untrained pattern which is proposed in work [14] for high – order TV-FRGs in testing phase.

For the training phase, we estimate all forecasted values for all high – order TV-FRGs based on fuzzy sets on the right-hand side within the same group. For each group in column 5 of Table 3, we divide each corresponding interval of each next state into p sub-intervals with equal length, and calculate a forecasted value for each group according to (6).

$$\mathrm{forecasted_{output}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}} subm_{kj} \tag{6}$$

where, ($1 \leq j \leq$n)

n is the total number of next states or the total number of fuzzy sets on the right-hand side within the same group.

$subm_{kj}$ is the midpoint of one of p sub-intervals (means the midpoint of ^{k}th sub-interval in which the historical data belong to this sub-interval, $1 \leq k \leq$ p) corresponding to j-th fuzzy set on the right-hand side, where the highest level of A_{kj} takes place in this interval.

For the testing phase, based on the master vote scheme, we calculate forecasted value for a group which contains the unknown linguistic value of the next state according to Eq. (7); where the symbol w_h means the highest votes predefined by user, m is the order of the fuzzy logical relationship, the symbols M_{t1} and M_{ti} denote the midpoints of the corresponding intervals of the latest past and other past linguistic values in the current state. From Table 3, it can be shown that group 20 has the fuzzy relationship $A_7, A_7, A_6 \rightarrow \#$ as it is created by the fuzzy relationship $\mathrm{F(1990), F(1991), F\ (1992) \rightarrow F(1993)}$; since the linguistic value of F (1993) is unknown within the historical data, and this unknown next state is denoted by the symbol '#'. Then, calculating value for "#" based on the current state of group G_{20} is determined by (7).

$$Forecated_{for\#} = \frac{(M_{t1}*w_h)+M_{t2}+\cdots+M_{ti}+\cdots+M_{tm}}{w_h+(\mathrm{m}-1)} \tag{7}$$

Based on the forecasted rules is created in (6) and (7), we complete forecasted results for all third-order FRGs are listed in Table 4.

From Table 4 and Table 1, we complete forecasted results for the enrolments of University of Alabama the period from 1971 to 1992 based on third-order fuzzy time series model with seven intervals are listed in Table 5.

Table 4. The complete forecasted values for all third – order FRGs.

Year	No group	Third-FRGs	Forecasted value
1974	G1	$A_1, A_1, A_1 \rightarrow A_2$	14625
1975	G2	$A_1, A_1, A_2 \rightarrow A_3$	15375
1992	G19	$A_4, A_7, A_7 \rightarrow A_6$	18875
1993	G20	$A_7, A_7, A_6 \rightarrow \#$	18667

Table 5. The complete forecasted outputs for enrollments based on the third-order FTS model.

Year	Actual	Fuzzified	Results
1971	13055	A_1	Not forecasted
1972	13563	A_1	Not forecasted
1973	13867	A_1	Not forecasted
1974	14696	A_2	14625
1992	18876	A_6	18875
1993	N/A	#	18667

The performance of proposed model is assessed with help of the root mean square error (RMSE) to compare the difference between the forecasted values and the actual values. The RMSE is calculated according to formula (8) as follows:

$$\mathrm{R}MSE = \sqrt{\frac{1}{n}\sum_{t=\alpha}^{n} (F_t - R_t)^2} \tag{8}$$

Where, R_t denotes actual value at year t, F_t is forecasted value at year t, n is number of the forecasted data, α is order of the fuzzy logical relationship**.**

4.2. Forecasting Model Combining High – Order FTS and PSO

The goal of this subsection is that we present the hybrid forecasting model by combining PSO algorithm with the high – order FTS model in Subsection 4.1 for adjusting the length each of intervals in the universe of discourse without increasing the number of intervals by minimizing the MSE value (8). The detailed descriptions of the hybrid forecasting model are given as follows:

In our model, each particle exploits the intervals in the universe of discourse of historical data Y (t). Let the number of the intervals be n, the lower bound and the upper bound of the universe of discourse U on historical data Y (t) be b_0 and b_n, respectively. Each particle is a vector consisting of n-1 elements b_i where $1 \leq i \leq n-1$ and $b_i \leq$

b_{i+1}. Based on these n-1 elements, define the n intervals as $u_1 = [b_0, b_1]$, $u_2 = [b_1, b_2]$,,..., $u_i = [b_{i-1}, b_i]$,… and $u_n = [b_{n-1}, b_n]$, respectively. When a particle moves to a new position, the elements of the corresponding new vector need to be sorted to ensure that each element bi ($1 \leq i \leq n-1$) arranges in an ascending order. The complete steps of the hybrid forecasting model are presented in Algorithm 1.

Algorithm 1: FTS-PSO algorithm

Input: historical time series data

Output: The forecasting results and the MSE value (Gbest=min (pbest))

1. Initialize:

Define the universe of discourse *U*

Generate random population of N particles including (number of order, number of interval, positions X_{id} and velocities V_{id} of all P_n particles)

Initialize the value of the weight factor *w*;

2. While the stop condition (maximum iterations or minimum MSE criteria) is not satisfied do

2.1. for particle id, ($1 \leq i \leq P_n$) do

i. Define fuzzy sets based on the current position of particle id

ii. Fuzzify all historical data by Step 4 in Subsection 4.1

iii. Create all $\alpha-$ order fuzzy logical relationships by Step 5 in Subsection 4.1

iv. Construct all $\alpha-$ order fuzzy relationship groups by Step 6 in Subsection 4.1

v. Calculate and defuzzification forecasting outputs by Step 7 in Subsection 4.1

vi. Compute the MSE values for particle id based on Eq. (8)

vii. Update the personal best position (pbest) of particle id according to the MSE values mentioned above.

end for

viii. Update the global best position (Gbest) of all particles according to the MSE values mentioned above.

4. for particle id, ($1 \leq i \leq p_n$) do

ix. move particle id to another position according to (1) and (2)

end for

x. update ω according to Eq. (3)

end while

End.

5. Experimental Results

The performance of the proposed model is evaluated based on two different datasets consisting of enrolments data of University of Alabama [2] the period from 1971 to 1992 which is shown in Figure 1 (a) and dataset of gasonline price in VietNam covering the period from January 4, 2018 through to July 2, 2019 which is shown in Figure 1 (b).

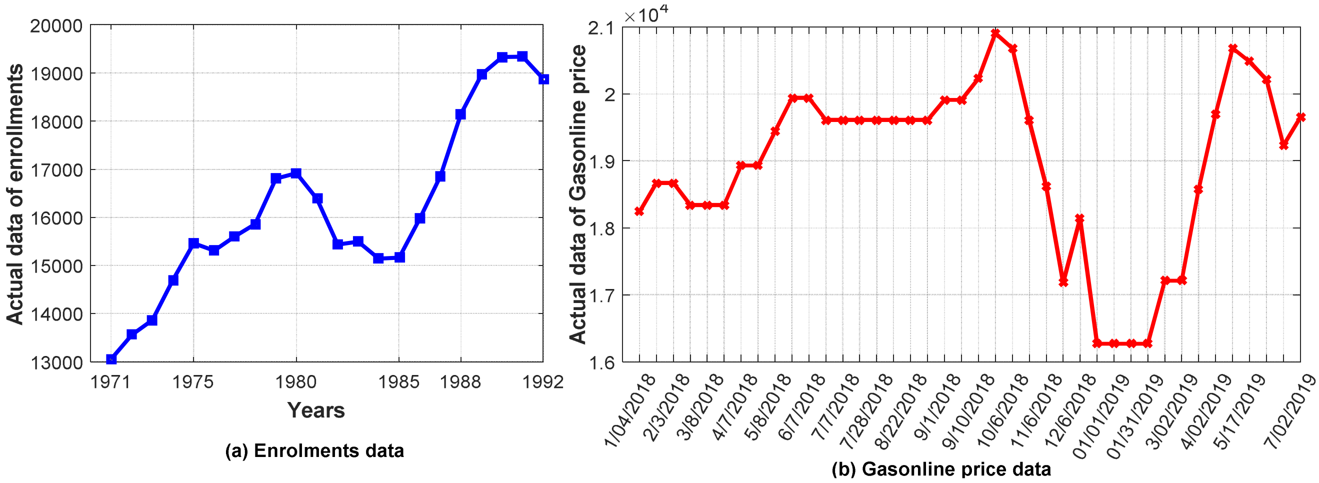

Figure 1. Historical data of enrollments of the University of Alabama and data of gasonline price in VietNam.

5.1. Forecasting the Enrollments

In this subsection, the proposed model is also applied to forecast the enrollments data of University of Alabama [2] from 1971 to 1992 and there is made a comparison of the forecasting results with the previous works [5, 10, 14, 29, 30].

A comparison of the forecasting results using RMSE (10) is shown in Table 6. From Table 6, it is obvious that proposed model has a MSE value is 44.99 which is the lowest among all forecasting models compared. The main difference among all the compared models is the fuzzy logical relationship group algorithms used to forecast. Five fuzzy forecasting models present in article [5, 10, 14, 29, 30] used the TV-FRGs algorithm to defuzzify forecasting output, while this study has proposed a method that benefits from the time-variant fuzzy relationship groups algorithm. As shown in Table 6, three of models; the HPSO, AFPSO and our model all use the PSO algorithm, but our proposed model gets smaller MSE values in forecasting. The following figure show the performance of the enrolments forecasting. The convergence of best objective values (MSE) and computational time for PSO based on the 5^{th}-order fuzzy time series with number of interval is 14 which are depicted in Figure 2.

In addition, we also perform 20 more runs to be compared with various high-order forecasting models under seven intervals including the models are introduced in papers [3, 5, 14, 10]. The detail of comparison is shown in Table 7. The trend in forecasting of enrollments based on the high-order FTS under various orders by the MSE value can be visualized in Figure 3.

Table 6. A comparison of the forecasted results of the proposed model with its counterparts based on different high – order FTS.

Year	Actual data	[30]	[29]	[5]	[14]	[10]	Proposed model
1971	13055	N/A	N/A	N/A	N/A	N/A	N/A
1972	13563	N/A	N/A	N/A	N/A	N/A	N/A
1973	13867	13813	13845	N/A	N/A	N/A	N/A
1974	14696	14681	14729	N/A	N/A	N/A	N/A
1989	18970	18926	18990	18910	18971	18974	18975.6
1990	19328	19275	19338	19334	19337	19338	19330.9
1991	19337	19428	19346	19334	19337	19335	19330.9
1992	18876	19046	18822	18910	18882	18882	18871.7
MSE		6825	1121	1101	234	173	44.99

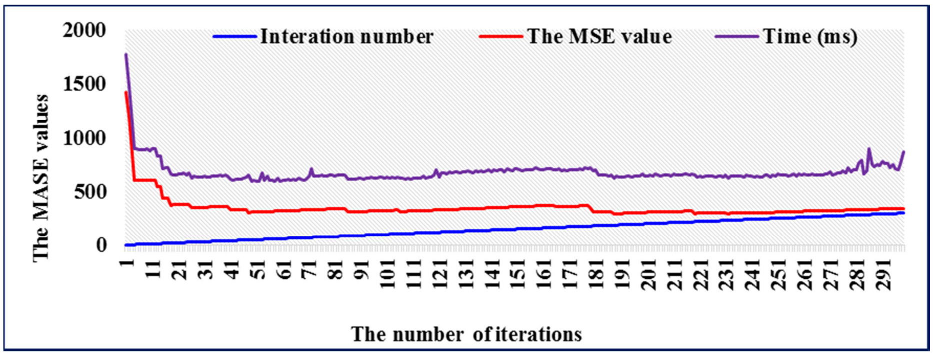

Figure 2. Convergence Speed and computational time of the proposed model using PSO.

Table 7. A comparison of the MSE value between proposed model and its counterparts under different number of orders and the number of interval equal to 7.

Models	2nd	3rd	4th	5th	6th	7th	8th	9th	Average
[3]	89093	86694	89376	94539	98215	104056	102179	102789	95868
[5]	67834	31123	32009	24984	26980	26969	22387	18734	31373
[14]	67123	31644	23271	23534	23671	20651	17106	17971	28121
[10]	19594	31189	20155	20366	22276	18482	14778	15251	20261
Our model	8451	1200	872	657	643	623	396	586	1678.5

Table 8. The forecasting results of the proposed model for gasonline in Viet Nam based on different high-orders.

Date	Actual data	2nd-order	3rd-order	4th-order	5th-order	6th-order	7th-order
1/4/2018	18240	-	-	-	-	-	-
1/19/2018	18670	-	-	-	-	-	-
2/3/2018	18670	18688.3	-	-	-	-	-
2/21/2018	18340	18322.6	18326	-	-	-	-
6/17/2019	19230	19260.3	19188.8	19176.6	19205.3	19209.5	19243.2
7/02/2019	19650	19639.3	19634.8	19709.8	19673.6	19640.7	19663.1
7/03/2019	N/A	19672	19666.3	19826.9	19977.4	19971.5	19683.7
MSE		12444.03	6428.6	3763	1084.9	630.93	583.23

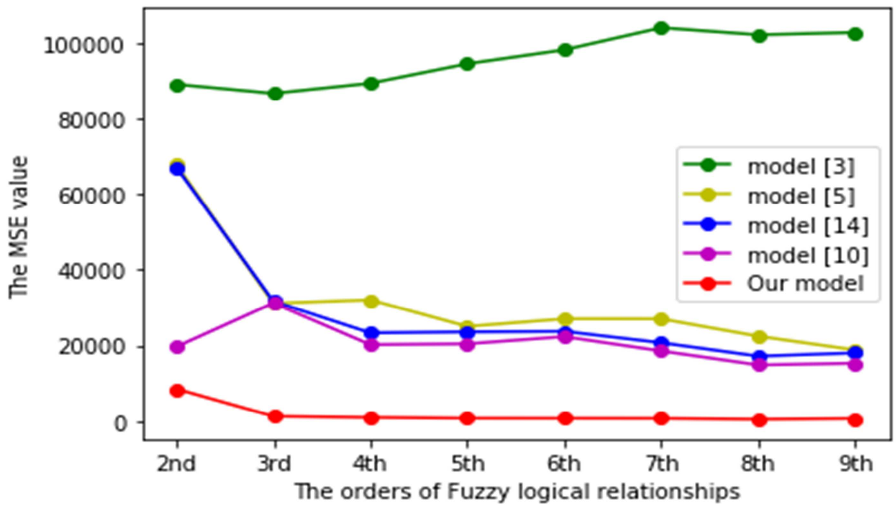

Figure 3. A comparison of the MSE values between the proposed model and its counterparts based on various high-order FTS.

During the simulation, the number of intervals is kept (the number of intervals=7) with different high – order FTS for the existing models and our model. A comparing of MSE value is listed in Table 7. From Table 7, it can be seen that the accuracy of the proposed model is improved significantly. Particularly, our model gets the lowest MSE value of 396 with 8th-order fuzzy relations and the average MSE value of the proposed model is 1678.5, which is smallest among six forecasting models compared.

5.2. Forecasting Gasonline Price in Vietnam

In this subsection, we apply the proposed model for forecasting the gas online price in Viet Nam which is presented in Figure 1 (b). This dataset taken from the site https://www.gso.gov.vn/Default.aspx?tabid=217. The forecasted accuracy of the proposed model is estimated using the MSE (8). To verify the superiority of the proposed model under various high-order FRGs, During simulation, the number of intervals is kept fix (number of intervals=14). The forecasting performance of proposed model is shown in Table 8. As shown in Table 8, the proposed model gets the smallest forecasting error rate by the MSE value of 583.23 based on 7^{th}-order for forecasting Gasonline in Viet Nam. Also form Table 8, it can be seen that the forecasting trend for gas online of the next day is gradually increasing.

6. Conclusion

In this paper, a new refined FTS forecasting model which combines the high – order FTS and PSO is presented. To reconcile the drawbacks of the conventional high-order FTS model which use the fuzzy relation groups, the time-variant high-order FRGs is used in this paper. By adopting the time-variant high-order FRGs which helps a more effective use of the historical data and has been proved to be more suitable for practical use. The paper also proposes a new defuzzification method to calculate the forecasted output values, which has been the main contribution issue to propose the forecasting model in this paper. In addition, the PSO are used to get the optimal partition of the interval in the universe of discourse. Among the heuristic optimization technique, the comparative results show the PSO method was generally found to perform better than other algorithms in terms of success rate and solution quality. By combining the TV-FRGs with PSO, the performance of the proposed forecasting model can be improved significantly. From the empirical study for forecasting enrolments and Gasonline price, the experimental results show that the proposed model outperforms its counterparts based on high – order FTS with various orders and different interval lengths. The detail of comparison was presented in Tables 5-7 and Figures 2-3. These results are very promising for the future work on the development of fuzzy time series in real-world forecasting applications.

Acknowledgements

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

References

[1] Bose1, & Mali, K. A novel data partitioning and rule selection technique for modelling high – order fuzzy time series. Applied Soft Computing, 2017. https://doi.org/10.1016/j.asoc.2017.11.011.

[2] Chen, S. M., 1996. Forecasting Enrolments based on Fuzzy Time Series, Fuzzy set and systems, vol. 81, 311-319, 1996.

[3] Chen, S. M., 2002. Forecasting enrolments based on high-order fuzzy time series, Cybernetics and Systems, vol. 33, no. 1, 1-16, 2002.

[4] Chen, S. M., Bui Dang, H. P. Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors. Knowledge-Based Systems 118, 204–216, 2017.

[5] Chen, S. M., Chung, N. Y., 2006b. Forecasting enrolments using high-order fuzzy time series and genetic algorithms. International Journal of Intelligent Systems, 21, 485–501, 2006b.

[6] Chen, S. M., Tanuwijaya, K. Fuzzy forecasting based on high-order fuzzy logical relationships and automatic clustering techniques, Expert Systems with Applications 38, 15425–15437, 2011.

[7] Chen, S. M., Jian, W. S. Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups, similarity measures and PSO techniques. Information Sciences 391–392, 65–79, 2017.

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

[9] Egrioglu, E., Aladag. C. H., Yolcu. Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks. Expert Systems with. Applications, 40 (3), 854–857, 2013.

[10] Huang, Y. L., et al. A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization, Expert Systems with Applications 7 (38), 8014–8023, 2011.

[11] Huarng, K. Effective lengths of intervals to improve forecasting in fuzzy time series, Fuzzy Sets and Systems, vol. 123, no. 3, 387-394, 2001.

[12] Hwang, J. R., Chen, S. M., & Lee, C. H. Handling forecasting problems using fuzzy time series. Fuzzy Sets and Systems, 100, 217–228, 1998.

[13] Kennedy, J., Eberhart, R. Particle swarm optimization, in: Proceedings of the 1995 IEEE International Conference on Neural Networks, 4, Perth, WA, Australia, 1942–1948, 1995.

[14] Kuo, I-H., et al. An improved method for forecasting enrolments based on fuzzy time series and particle swarm optimization, Expert systems with applications, 36, 6108–6117, 2009.

[15] Kuo, I-H., et al. Forecasting TAIFEX based on fuzzy time series and particle swarm optimization, Expert Systems with Applications 2 (37), 1494–1502, 2010.

[16] Lee, L. W., Wang, L. H., Chen, S. M. Temperature prediction and TAIFEX forecasting based on high order fuzzy logical relationship and genetic simulated annealing techniques, Expert Systems with Applications, 34, 328–336, 2008.

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

[18] Ling-Yuan Hsu et al,. Temperature prediction and TAIFEX forecasting based on fuzzy relationships and MTPSO techniques, Expert Systems with Applications. 37, 2756–2770, 2010.

[19] Nguyen Cong Dieu, Nghiem Van Tinh. Fuzzy time series forecasting based on time-depending fuzzy relationship groups and particle swarm optimization, Proceedings of the 9th National conference on Fundamental and Applied Information Technology Research (FAIR'9), pp. 125-133, 2016.

[20] Nghiem Van Tinh, Nguyen Cong Dieu. An improved method for stock market forecasting combining high-order timevariant fuzzy logical relationship groups and particle swam optimization in: Proceedings of the International Conference, ICTA, pp. 153-166, 2016.

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

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

[23] Song, Q., Chissom, B. S., Forecasting enrolments with fuzzy time series-Part I, Fuzzy Sets and Systems, vol. 54, no. 1, 1-9, 1993a.

[24] Song, Q., Chissom, B. S. Fuzzy time series and its models, Fuzzy Sets and Systems, vol. 54, no. 3, 269-277, 1993b.

[25] 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. Lecture Note in Electrical Enginearing 406, 181-189, Springer, 2016.

[26] Tinh N. V., Dieu N. C. A novel forecasting model based on combining time-variant fuzzy logical relationship groups and K-means clustering technique. Proceedings of the 9th National Conference on Fundamental and Applied Information Technology Research (FAIR10), 2017, DOI: 10.15625/vap.2017.0002.

[27] Yu, H. K., A refined fuzzy time-series model for forecasting, Physical A: Statistical Mechanics and Its Applications, vol. 346, no. 3-4, 657-681, 2005.

[28] Yu, H. K., Weighted fuzzy time series models for TAIFEX forecasting, Phys. A, Stat. Mech. Appl. 349 (3–4), 609–624, 2005.

[29] Wang Hongxu, Guo Jianchun, Feng Hao, and Jin Hailong, "An improved forecasting model of fuzzy time series," Applied Mechanics and Materials. vol. 678, pp. 64–69, [3rd International Conference on Mechatronics and Control Engineering (ICMCE 2014). 2014].

[30] Wang HongXu, JianChun Guo, Hao Feng, and HaiLong Jin, "A fuzzy time series forecasting model based on data differences," 1 ICT IN EDUCATION. Frontiers in Computer Education. pp. 15–18, 2014 [2nd International Conference on Frontiers in Computer Education (ICFCE2014). Wuhan, China, 2014].