https://doi.org/10.48048/tis.2022.2157

Developing a Fuzzy Time Series Forecasting Model Based on Hedge Algebras and Particle Swarm Optimization

Nghiem Van Tinh*, Nguyen Tien Duy and Tran Thi Thanh

Faculty of Electronics, Thai Nguyen University of Technology, Thai Nguyen, Vietnam

(*Corresponding author's e-mail: nghiemvantinh@tnut.edu.vn)

Received: 12 September 2020, Revised: 5 May 2021, Accepted: 12 May 2021

Abstract

In recent years, numerous fuzzy time series (FTS) forecasting models have been widely used. One of the important factors for obtaining high forecasting accuracy in fuzzy time series model is that the lengths of intervals in the universe of discourse. In this study, a hybrid forecasting model which uses hedge algebra (HA) and particle swarm optimization (PSO) is proposed to determine optimal lengths of intervals in FTS models. In that, HA is utilized as a tool to partition the universe of discourse into intervals with unequal-size corresponding to the semantic intervals calculated from the linguistic terms. After processing of generating the intervals, we define fuzzy sets based on the observation data of times series and use them to establish fuzzy relationship groups. Then, the proposed model is combined with the PSO technique to find the appropriate length of each interval with view to reaching the better forecasting accuracy rate. The performance of the proposed model is evaluated with the historical data of enrolments at the University of Alabama. The simulated results obtained indicate that the proposed model achieves higher forecasting accuracy compared other existing forecasting models and it can obtain better quality solutions for both the 1st-order and high-order FTS model.

Keywords: Fuzzy time series, Fuzzy relationship group, Hedge algebras, Particle swam optimization, Enrolments

Introduction

Dealing with the time series forecasting problem, many forecasting models have been introduced to advance the decision-making process concerning future, such as enrollments forecast for the next year, temperature prediction of the coming days, annual population forecasting, financial forecasting, ... Based on fuzzy set theory [1], Song and Chissom [2,3] proposed 2 FTS models by using max-min operations in fuzzy relationships to forecast the enrollments of the University of Alabama. Compared with the previous traditional forecasting models, such as regression analysis, moving average, autoregressive moving average and ARIMA model, the forecasting models in articles [2,3] can make better predictions with forecasting problems in which the historical data needs to be represented by linguistic values or uncertain data series. However, their models 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, to avoid this shortcoming, Chen [4] developed a FTS forecasting model using simplified arithmetic operations rather than max-min composition operator in defuzzification process. In addition, research works in articles [5,6] pointed out the importance of assigning weights to resolve the issue of recurrent fuzzy relationship and to reflect the difference in their importance. From the expansion of the research [4] into a high-order fuzzy time series model [7] and the influence of the lengths of intervals in article [8] together with the development from the 1-factor FTS models into 2-factor FTS model [9] is the foundation for the strong development of FTS models in the next time periods. Recently, many authors have used different techniques in each stage of FTS model to improve forecasting accuracy. Chen and Tanuwijaya [10] used the automatic clustering method to partition the universe of discourse into different interval lengths in the fuzzification stage of the forecasting model. Some other researches combine optimization techniques with different FTS models to adjust and find the lengths of intervals from the universe of discourse [11-22]. Based on the idea of finding the suitable interval lengths, many models used clustering techniques to divide time series dataset into clusters, then adjust these clusters into intervals with different lengths such as: K-mean cluster [23,24] and clustered C-mean [25,26]. A completely different way from fuzzy approach, information granules consider to be a sound option. Just recently, several related works have been presented. Ho *et al*. [27] has introduced a forecasting model based on the theory of HA [28] to apply for forecasting university enrolments. In which, the HA was used to model linguistic domains and variables instead of performing data fuzzification and defuzzification. In addition, Tung *et al*. [29] proposed a HA-based forecasting model to find different lengths of intervals in the universe of discourse by mapping the semantics of linguistic terms into fuzziness intervals. However, 2 these research works only focus on building a 1st-order forecasting model to forecasting the number of enrollments at the University of Alabama.

From analyzing of the research works above showed that the lengths of intervals and the order of fuzzy relationships are 2 critical factors for forecasting accuracy. Bearing in mind the idea in the using HA, in this paper we propose a hybrid fuzzy time series model combining HA and PSO for forecasting the enrollments of the University of Alabama [4]. In this research, HA is used to partition the universe of discourse into unequal-sized intervals by quantifying the linguistic terms themselves which is used to describe the historical values of fuzzy time series. After generating the intervals, the historical time series dataset is fuzzified based on the defined fuzzy sets. Each fuzzified time series value is then used to create the FLRs and divide them into groups. Later, all these fuzzy relationship groups are utilized to obtain the forecasting results based on the our defuzzification principle [30]. Finally, the proposed model is combined with the PSO algorithm to adjust the initial interval lengths for further increasing predictive accuracy.

Materials and methods

In this section, we briefly review basic concepts related to fuzzy time series [2,3], the HA [28] and PSO algorithm [31].

Some basic definitions of FTS

Based on the fuzzy set theory [1], Song and Chissom [2,3] introduced the definition of FTS and constructed its model by means of fuzzy relational equations, in which the values of historical time series data are presented by fuzzy sets. Let $U = \{u_1, u_2, \ldots, u_n\}$ be an universe of discourse; a fuzzy set A of U can be defined as $A = \{\mu_A(u_1)/u_1 +, \ \mu_A(u_2)/u_2 \ldots + \mu_A(u_n)/u_n\}$, where $\mu_A : U \to [0,1]$ is the membership function of A, $\mu_A(u_i)$ indicates the degree of membership of u_i in the fuzzy set A, $f_A(u_i) \in [0, 1]$, and $1 \leq i \leq n$. The basic definitions related to FTS are summarized as below:

Definition 1: Fuzzy time series [2,3]
Let $Y(t)$ $(t = .., 0, 1, 2 ..)$, a subset of real numbers, be the universe of discourse on which the fuzzy sets $f_i(t)$ $(i = 1, 2 \ldots)$ are defined in the universe of discourse $Y(t)$ and $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$, then $F(t)$ is called a FTS definition on $Y(t)$ (t ..., 0, 1, 2 ...).

Definition 2: Fuzzy logical relationship - FLR [2-4]
If there exists a fuzzy relationship R(t-1,t), such that F(t) = F(t-1)$*$ R(t-1,t), where "$*$" is an max-min composition operator, then F(t) is said to be caused by F(t-1). The relationship between F(t) and F(t-1) can be denoted by F(t-1) $\to$ F(t). Let A_i = F(t) and A_j = F(t-1), the relationship between F(t) and F(t-1) is denoted by fuzzy logical relationship $A_i \to A_j$ where A_i and A_j refer to the current state or the left-hand side and the next state or the right-hand side of fuzzy relationship.

Definition 3: λ-order fuzzy logical relationship [7]
Let F(t)be a fuzzy time series. If F(t) is caused by F(t-1), F(t-2),…, F(t-λ+1), F(t-λ) then this fuzzy relationship is represented by F(t-λ), …, F(t-2), F(t-1)$\to$ F(t) and is called an $\boldsymbol{\lambda}$-order fuzzy time series.

Definition 4: Time-variant fuzzy relationship groups (TV-FRGs) [17]
The fuzzy logical relationship is defined by the relationship F(t-1)$\to$ F(t). If, let $F(t) = A_i(t)$ and $F(t-1) = A_j(t-1)$. The FLR between F(t-1) and F(t) can be denoted as $A_j(t-1) \to A_i(t)$. Also at the time t, we have the following fuzzy logical relationships: $A_j(t1-1) \to A_{i1}(t1), .., A_j(t\lambda - 1) \to A_{i\lambda}(t\lambda)$ with $t1, t2, .., t\lambda \leq t$. It is noted that $A_i(t1)$, $A_i(t1)$,…, and $A_i(t\lambda)$ with the same fuzzy set A_i but appear at different times t1, t2,…, and tn, respectively. It means that if these FLRs occur before $A_j(t-1) \to A_i(t)$, we can group these FLRs into a FRG according to the left-hand side of each FLR as $A_j(t-1) \to A_{i1}(t1), A_{i2}(t2), \ldots, A_{i\lambda}(t\lambda), A_i(t)$. It is named 1st- order TV-FRGs.

Some basis concepts of hedge algebras [28]

HA are created by Ho and Wechler [28] in 1990. In the field of time series forecasting, this theory is considered as a new approach to quantify the linguistic terms differing from the fuzzy set approach. In this study, the concepts of HA are employed as basis to partition the universe of discourse of time series into initial intervals with different lengths. Assume that there is a set of linguistic values of linguistic variable X which are sorted as follows: $X = \{\text{Very Very low} < \text{Very low} < \text{low} < \text{Little low} < \text{Very Little low} < \text{medium} < \text{Very Little high} < \text{Little big} < \text{high} < \text{Very high} < \cdots\}$. Each of linguistic variable $\mathcal{X}$ is represented by an algebraic structure as $\mathcal{AX} = (X, G, C, H, \leq)$ and called HA, where X is the set of terms in $\mathcal{X}$; $\leq$ denotes a natural semantically ordering relation on X; $G = \{c^-, c^+\}$, $c^- \leq c^+$, is the set of primary generators, in which c^+ and c^- are, respectively, the negative primary term and the positive one of a linguistic variable X, C = {0,1,w} a set of constants, with ($\mathbf{0} \leq c^- \leq \mathbf{W} \leq c^+ \leq \mathbf{1}$); $H = H^- \cup H^+$, với $H^- = \{h_{-q} \geq \cdots \geq h_{-2} \geq h_{-1}\}$ is the set of all negative hedges of X, $\forall h \in H^-$ then $hc^+ \leq c^+$ and $H^+ = \{h_1 \leq h_2 \leq \cdots \leq h_p\}$ is the set of all positive ones of X, $\forall h \in H^+$ then $hc^+ \geq c^+$. Example $H^- = \{\text{Little} > \text{Rather}\}$, $H^+ = \{\text{More} < \text{Very}\}$. $\forall x \in X$, $x = h_n h_{n-1} \ldots h_1 c$, $h_j \in H$ with $c \in G$. If X and H are linearly ordered sets, then $\mathcal{AX} = (X, G, C, H, \leq)$ is called linear HA, furthermore, if AX is equipped with additional operations $\sum$ and Φ that are, respectively, infimum and supremum of H(x), then it is called complete linear hedge algebras (ClinHA) and denoted $\mathcal{AX} = (X, G, C, H, \sum, \Phi \leq)$ [32]. Some general definitions of HA are given as follows:

Definition 5: Let AX = (X, G, C, H, ≤) be a ClinHA. $fm: X \to [0, 1]$ is said to be a fuzziness measure of terms in X if:

1) $fm(c^-) + fm(c^+) = 1$ and $\sum_{h \in H} fm(hx) = fm(x)$, with $\forall x \in X$.
2) For the constants **0**, **W** and **1**, $fm(\mathbf{0}) = fm(\mathbf{W}) = fm(\mathbf{1}) = 0$.
3) For $\forall x, y \in X, \forall h \in H$, $\frac{fm(hx)}{fm(x)} = \frac{fm(hy)}{fm(y)}$, that is this proportion does not depend on specific elements and, therefore, it is called fuzziness measure of the hedge h and denoted by $\mu(h)$. The properties of $fm(x)$ and $\mu(h)$ are introduced as follows:

Proposition 1: Let fm is the fuzziness measure function on X, the following statements hold. With $\in X, x = h_n h_{n-1} \ldots h_1 c, h_j \in H, c \in G$

1) $fm(hx) = \mu(h) fm(x), \forall x \in X$.
2) $\sum_{-q<i<p, i\neq 0} fm(h_i c) = fm(c)$.
3) $\sum_{-q<i<p, i\neq 0} fm(h_i x) = fm(x)$.
4) $fm(x) = fm(h_n h_{n-1} \ldots h_1 c) = \mu(h_n)\mu(h_{n-1}) \ldots \mu(h_1) fm(c)$.
5) $\sum_{i=-1}^{-q} \mu(h_i) = \alpha$ and $\sum_{i=1}^{p} \mu(h_i) = \beta$, with $\alpha, \beta > 0$ and $\alpha + \beta = 1$.

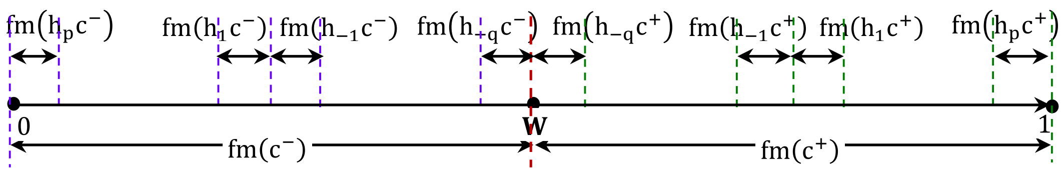

Figure 1 The order of elements $x \in X, h_j \in H, c \in G$.

Definition 6: The fuzziness interval of the linguistic terms x ∈ X, denoted by $\Im(x)$, is a subinterval of [0,1], if $|\Im(x)| = fm(x)$ where $|\Im(x)|$ is the length of fm(x), and recursively determined by the length of x as follows:

1) If length of x is equal to 1 (l(x) = 1), that mean $x \in \{c^-, c^+\}$, then $|\Im(c^-)| = fm(c^-)$, $|\Im(c^+)| = fm(c^+)$ and $\Im(\mathbf{c}^-) \leq \Im(c^+)$;

2) Suppose that n is the length of x (l(x) = n) and fuzziness interval $\Im(x)$ has been defined with $|\Im(x)| = fm(x)$. The set $\{\Im(h_j x) | j \in [-q^p]\}$, where $[-q^p] = \{j \mid -q \leq j \leq -1 \text{ or } 1 \leq j \leq p\}$, is a partition of $\Im(x)$ and we have: for $Sgn(h_p x) = -1$, $\Im(h_p x) \leq \Im(h_{p-1} x) \leq \ldots \leq \Im(h_1 x) \leq \Im(h_{-1} x) \leq \ldots \leq \Im(h_{-q} x)$; for $Sgn(h_p x) = +1$, $\Im(h_{-q} x) \leq \Im(h_{-q+1} x) \leq \ldots \leq \Im(h_{-1} x) \leq \Im(h_1 x) \leq \ldots \leq \Im(h_p x)$.

Particle swarm optimization algorithm

PSO algorithm had been proposed by Kannedy and Eberhart [31] in 1965. It is considered as a tool to resolve with optimization problems, where a set of potential solutions is represented by a swarm of particles and each particle is move through the search space (d-dimensional space) for search the optimal solution. When particles moving, all particles (i.e, N particles) have fitness values which are evaluated by fitness function and the position of the best particle among all particles found so far is kept and each particle keeps its personal best position which has passed previously. In the movement process of particles, each k^{th} ($1 \leq k \leq N$) particle associated with the velocity vector $V_{ki} = [v_{k,1}, v_{k,2}, \dots, v_{k,d}]$ and the position vector $X_{ki} = [x_{k,1}, x_{k,2}, \dots, x_{k,d}]$ of particle are updated by the best position $P_{best\_kd} = [p_{k,1}, p_{k,2}, \dots, p_{k,d}]$ encountered by the particle so far and the best position global $G_{best} = \min(P^t_{best\_kd})$ found by the overall best out of all the particles in the population. The briefly summarizes steps of the standard PSO algorithm in Algorithm 1 as follows:

Algorithm 1 The standard PSO algorithm

Step 1: Initialize random positions x_{ki} ; random velocities v_{ki} in d dimensional space (i = 1,2,…,d);

- Positions of each particle are randomly determined and saved in a vector X_{kd} as follows:

$$X_{ki} = [x_{k,1}, x_{k,2}, \dots, x_{k,d}] \tag{1}$$

Where; x_{ki} denotes i^{th} position of k^{th} particle. N is the number of particles in a swarm;

- Velocities are randomly determined and stored in a vector V_{kd} given below:

$$V_{ki} = [v_{k,1}, v_{k,2}, \dots, v_{k,d}] \tag{2}$$

Step 2: According to the fitness function value f(x), the values of P_{best_kd} and G_{best} of particles given in Eq. (3) are determined as: $P_{best\_kd} = [p_{k,1}, p_{k,2}, \dots, p_{k,d}]$.

Where, P_{best_kd} is a vector stores the positions corresponding to the k^{th} particle's best individual performance and calculated as:

$$P^{t+1}_{best\_kd}\, f(x) = \begin{cases} P^{t+1}_{best\_ki}, & if\ f(x^{t+1}_{ki}) > P^t_{best\_ki} \\ f(x^{t+1}_{ki}), & if\ f(x^{t+1}_{ki}) \leq P^t_{best\_ki} \end{cases} \tag{3}$$

and $G_{best} = \text{minimum}(P_{best\_ki})$ denotes the best one of all personal best positions of all particles within the swarm.

Step 3: c_1 and c_2 are 2 learning factors which control the influence of the social and cognitive components ($c_1 = c_2 = 2$), respectively. ω is the time-varying inertia weight.

In each iteration t, the parameter ω is calculated as follows:

$$\omega^t = \omega_{max} - \frac{t * (\omega_{max} - \omega_{min})}{iter\_max} \tag{4}$$

Where, $iter\_max$ denotes the maximum iteration number.

Step 4: The new velocity and position of each particle can be updated by using the current velocity and distance from the P_{best} to G_{best} as shown in Eqs. (4) and (5), respectively.

$$V^{t+1}_{ki} = \omega^t * V^t_{ki} + c_1 * R1(\,) * \left(P_{best\_kd} - X^t_{ki}\right) + c_2 * R2(\,) * (G_{best} - X^t_{ki}) \tag{5}$$

$$X^{t+1}_{ki} = X^t_{ki} + V^{t+1}_{ki} \tag{6}$$

Where, R1 and R2 are generated random values in the domain [0,1].

Step 5: Steps 2 to 4 are repeated until a predetermined maximum iteration number ($iter\_max$) is reached.

A hybrid FTS forecasting model based on HA and PSO

This section, we introduce a novel FTS forecasting model which is combined between HA and PSO for improving forecasted accuracy. The framework of proposed model is presented in **Figure 2**, including 3 stages; the 1st stage is to partition the historical data into initial intervals based on HA; the 2nd stage is to establish the FTS model and create forecasting rules and the final stage is to find optimal lengths of intervals by applying PSO algorithm. To implement these stages, all historical enrollments data [4] are utilized for illustrating forecasting process. This dataset has been selected to forecast with the great amount of research works which have been published in the literatures [3,4,6-9,10-12,14,15,17,21-24,27,29,30]. The 3 stages of proposed model are present as follows.

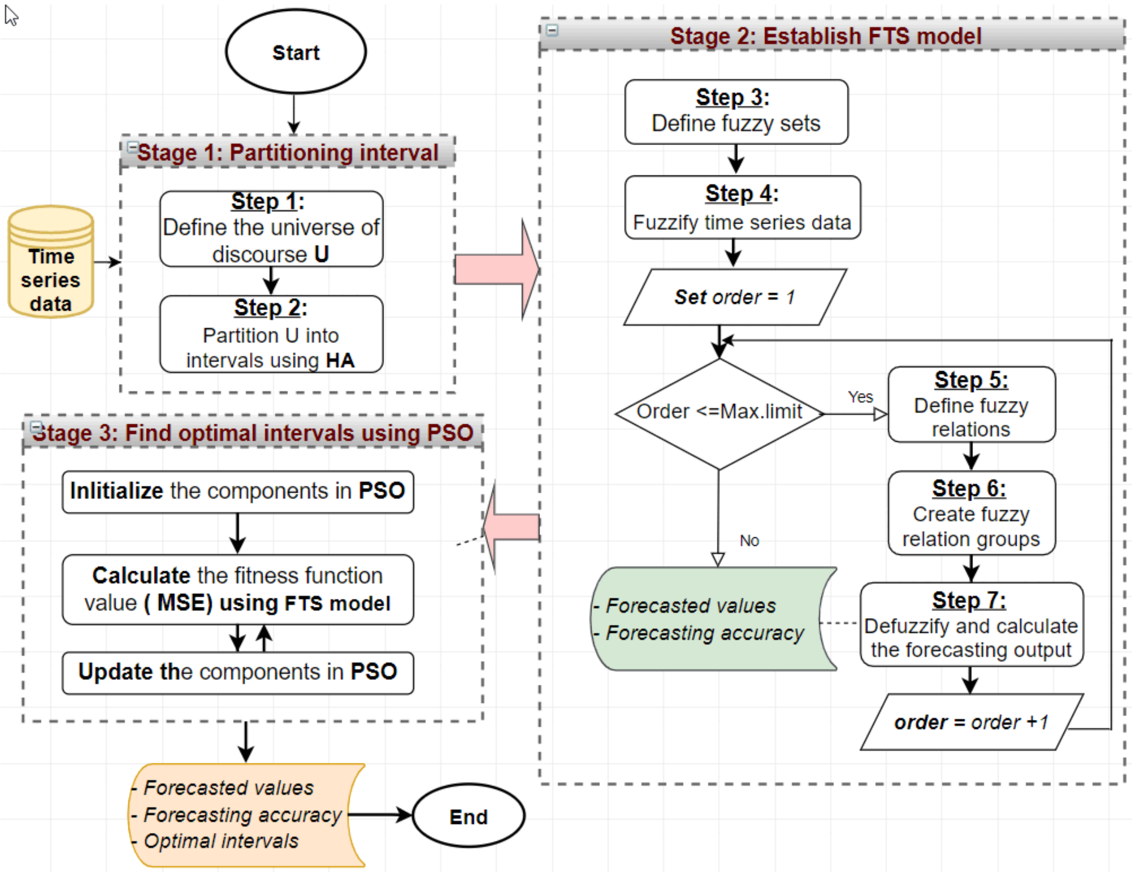

Figure 2 flowchart of the proposed FTS forecasting model.

Stages 1 and 2: Constructing forecasting model based on FTS and HA

In this section, a forecasting model based on combining the FTS and HA for forecasting enrolments of the University. Initially, the HA is applied to divide the universe of discourse into initial intervals with unequal-lengths by quantitative mapping of linguistic terms into fuzzy intervals. Based on these newly obtained intervals, we defined fuzzy sets and fuzzy historical data on each divided interval. From these fuzzified values, we derive the FLRs and establish fuzzy relationship groups according to [17]. Later, all these FRGs are used to obtain the forecasting results based on the defuzzification method [30]. The proposed forecasting model can be given step-by-step as follows:

Step 1: Define the universe of discourse U of historical time series data

Let $U = [D_{min} - D_1, D_{max} + D_2]$ is universe of discourse. To define U, the minimum value D_{min} and the maximum value D_{max} of the historical time series data is defined. In order to ensure the forecasting values bounded in the universe of discourse U 2 positive integers D_1 and D_2 are properly selected, respectively. From historical enrolments time series, U is defined as U = [13000, 20000], where D_{min}, = 13055, D_{max} = 19337, D_1 = 55, D_2 = 663, LU = 7000.

Step 2: Partition U into different intervals based on HA

This step uses HA with structure as $\mathcal{AX} = (X, G, C, H, \leq)$, where X is the set of terms of the linguistic variable "enrollments"{X = dom(enrollments)}; $\leq$ denotes a natural semantically ordering relation on X; $G = \{c^-, c^+\}$ = {Low, High}, Low (Lw) ≤ High (Hi); C = {0, w, 1} a set of constants, with ($0 \leq c^- \leq W \leq c^+ \leq 1$) and H = {Very, Little}. To compare the forecasted results of the proposed model with other models. In this paper, we use the number of intervals equal to 7 and 14 which are the number of linguistic terms used to quantify the time series values. In particular, suppose that the linguistic terms are given in **Table 1**. Based on these linguistic terms, we define the corresponding intervals of them in domain U.

Table 1 The number of language terms.

Number of linguistic terms	Linguistic terms and its order
7	A_1 = Very Very Low (VVLw) < A_2 = Little Verry Low (LVLw) < A_3 = Little Little Low (LLLw) < A_4 = Very Little Low (VLLw) < A_5 = Very Little High (VLHi) < A_6 = Little Little High (LLHi) < A_7 = Very High (VHi)
14	A_1 = VVLw < A_2 = LLVLw < A_3 = VLVLw < A_4 = VLLLw< A_5 = LLLLw < A_6 = LVLLw < A_7 = VVLLw < A_8 = VVLHi < A_9 = LVLHi < A_{10} = LLLHi < A_{11} = VLLHi < A_{12} = VLVHi <A_{13} = LLVHi < A_{14} = VVHi

Step 2.1: The domain U = [13000, 20000] is mapped to the domain [0,1]

Suppose the value of 16807 in the time series dataset is the average value, then we can set up the following parameters: fm(Low) = $\frac{16807 - 13000}{20000 - 13000} = 0.544$, fm(High) = 1 - 0.544 = 0.456 and LU = 20000 – 13000 = 7000. Mapping these values to U, we have covfm(Low) and covfm(High) that are determined, respectively as fm(Low)×LU = 0.544×7000 = 3808, fm(high)×LU = 0.456×7000 = 3192. In this paper, we can choose μ(Little) = 0.48, μ(Very) = 1 - 0.48 = 0.52. Based on μ(Little), μ(Very) value, the value of α , β is defined as 0.48, 0.52, respectively.

From here, the fuzziness interval of linguistic terms in the domain [0,1] can be calculated: fm(VVLw) = 0.1471, fm(LVLw) = 0.1358, fm(LLLw) = 0.1253, fm(VLLw) = 0.1358, fm(VLHi) = 0.11138, fm(LLHi) = 0.1051, fm(VHi) = 0.2371.

Step 2.2: Define the fuzzy interval of linguistic variable in the universe of discourse

Based on Step 2.1, the linguistic values of terms belong to fuzziness interval is calculated as follows:
covfm(A_1) = μ(Verry)×μ(Very)×covfm(Low) = 0.52×0.52×3808 = 1029.683;
covfm(A_2) = μ(Little)×μ(Very)×covfm(Low) = 0.48×0.52×3808 = 950.477;
covfm(A_3) = μ(Little)×μ(Little)×covfm(Low) = 0.48×0.48×3808 = 479.36;
covfm(A_4) = μ(Very)×μ(Little)×covfm(Low) = 0.52×0.48×3808 = 950.477;
---;
covfm(A_7) = μ(Very)×covfm(High) = 0.52×3192 = 1659.84

Mapping the value of linguistic terms to the domain of the universe of discourse U, we get the intervals corresponding to linguistic terms, which are listed as below:

For 7 linguistic terms, obtaining 7 intervals as u_1 = [13000, 14029.68), u_2 = [14029.68, 14980.2), u_3 = [14980.2, 15,857.5), u_4 = [15857.5, 16808), u_5 = [16808, 17605), u_6 = [17605, 18340.16), u_7 = [18340.16, 20000].

For 14 linguistic terms, obtaining 14 intervals as u_1 = [13000, 13539.5), u_2 = [13539.5, 14079), u_3 = [14079, 14438.5), u_4 = [14438.5, 14798), u_5 = [14798, 15157.5), u_6 = [15157.5, 15517), u_7 = [15517, 15756.5), u_8 = [15756.5, 15996), u_9 = [15996, 16316.5), u_{10} = [16316.5, 16637), u_{11} = [16637, 17117.5), u_{12} = [17117.5, 17598), u_{13} = [17598, 18799), u_{14} = [18799, 20000].

Step 3: Define linguistic terms A_i which represented by fuzzy sets
Each of interval in Step 2 represents a linguistic value of linguistic variable "enrolments". For 7 intervals, there are 7 linguistic values to represent different regions in the universe of discourse on U. Each linguistic value represents a fuzzy set A_i and its definitions is described in Eqs. (7) and (8) as follows:

$$\mathrm{A_i} = {}^{a_{i1}}/_{\mathrm{u}_1} + {}^{a_{i2}}/_{\mathrm{u}_2} + \ldots + {}^{a_{ij}}/_{\mathrm{u}_j} + \cdots + {}^{a_{i7}}/_{\mathrm{u}_7} \tag{7}$$

$$a_{ij} = \begin{cases} 1 & j = i \\ 0.5 & j = i-1 \; or \; j = i+1 \\ 0 & otherwise \end{cases} \tag{8}$$

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 the universe of discourse. 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 triangular membership function and which is defined Eq. (8). From Eqs. (7) and (8), a fuzzy set contains 7 intervals. Contrarily, each of interval belongs to all fuzzy sets with different membership degrees. Therefore, based on the obtained intervals from enrollments dataset, the corresponding linguistic values are illustrated in **Figure 3**.

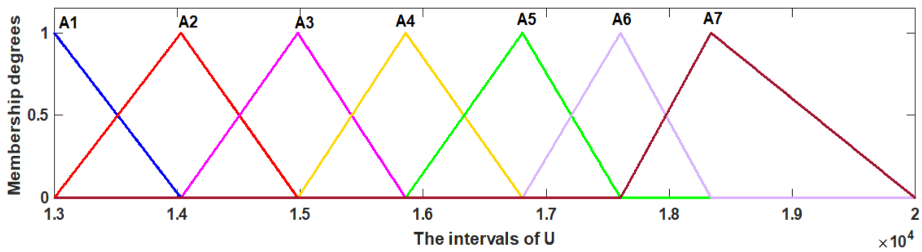

Figure 3 The fuzzy sets are defined by intervals using the triangular membership function.

Step 4: Fuzzy the historical time series data
To fuzzify the historical time series data, it is essential to obtain the degree of membership value of each data value belonging to each u_i for each year. If the maximum membership value of 1 day's observation occurs at u_i , and $(1 \le i \le 7)$, then the fuzzified value for that particular year is considered 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, 14078.56). So, we then assign the linguistic value ''Very Very Low'' (e.g., the fuzzy set A_1) corresponding to interval u_1 to it. Considering 2 time series 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. (7), 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 enrolments data are presented in **Table 2**.

Table 2 Fuzzified historical enrollments data of the University of Alabama.

Year	Real data	Fuzzy sets	Linguistic values
1971	13055	A_1	Very Very Low
1972	13563	A_1	Very Very Low
---	---	---	--------------
1991	19337	A_7	Little Little High
1992	18876	A_7	Little Little High

Step 5: Create all λ-order fuzzy logical relationships ($\lambda \geq 1$).
Based on Definitions 3 and 4, to establish a λ-order fuzzy logical relationship, we should find out any relationship which has the $F(t-\lambda), F(t-\lambda+1), \ldots, F(t-1) \rightarrow F(t)$, where $F(t-\lambda), F(t-\lambda+1), \ldots, F(t-1)$ and $F(t)$ are called the current state and the next state, respectively. Then a λ-order fuzzy relationship in the training phase is got by replacing the corresponding linguistic values. For example, supposed $\lambda = 1$ from **Table 2**, a fuzzy relation $A_1 \rightarrow A_1$ is got as $F(1971) \rightarrow F(1972)$. So on, we get the 1st-order fuzzy relationships are shown in column 4 of **Table 3**, where there are 22 fuzzy relationships; the first 21 relations are called the trained patterns, and the last one is called the untrained pattern (in the testing phase). For the untrained pattern, relation 22 has the fuzzy relation $A_7 \rightarrow \#$ as it is created by the relation $F(1992) \rightarrow F(1993)$, since the linguistic value of $F(1993)$ is unknown within the historical data, and this unknown next state is denoted by the symbol '#'. The same way, suppose $\lambda = 2$, a fuzzy relationship $(A_1, A_1) \rightarrow A_1$ is got as (F(1971), F(1972)) → F(1973). The complete 2-order fuzzy relationships are listed in column 5 of **Table 3**.

Table 3 The complete the 1st-order and the 2nd-order FLRs.

No	Year	Fuzzy set	1st-order FLRs	2nd-order FLRs
	1971	A_1		
1	1972	A_1	$A_1 \rightarrow A_1$	
2	1973	A_1	$A_1 \rightarrow A_1$	$A_1, A_1 \rightarrow A1$
3	1974	A_2	$A1 \rightarrow A_2$	$A_1, A_1 \rightarrow A_2$
---	---	---	---	---
21	1992	A7	$A_7 \rightarrow A_7$	$A_7, A_7 \rightarrow A_7$
22	1993	#	$A_7 \rightarrow \#$	$A_7, A_7 \rightarrow \#$

Step 6: Construct all fuzzy relationship groups (FRGs)
In this step, an approach is different from the approach in articles [4,10-14] in the way where the fuzzy relationship groups are created. In previous approaches [4,5], the recurrent FLRs were simply ignored when fuzzy relationship groups were established or these repeated fuzzy relationships is mentioned, but it is not suitable at each of forecasting time, respectively. In this study, we rely on a concept of time-variant fuzzy logical relationship group [17] which presented in Definition 5 to create FRGs, called TV-FRGs. To explain this, assume that $\lambda = 1$. We consider the 3 1st-order FLRs at 3 different times t = 1972, 1973, 1974 in column 4 of **Table 3** as follows: F(t = 1972): $A_1 \rightarrow A_1$; F(t = 1973): $A_1 \rightarrow A_1$; F(t = 1974): $A_1 \rightarrow A_2$; all of them with the same fuzzy set A_1 on the left – hand side. Then, if considering at forecasting time t = 1992, we have obtained a 1st-order FRG (i.e., No 1) as follows: $A_1 \rightarrow A_1$. With forecasting time t = 1993, before that there 2 FLRs with the same on left-hand side, these FLRs can be grouped into a FRG as No 2: $A_1 \rightarrow A_1, A_1$. With forecasting time t as 1994, then the group G3 is expressed as follows $A_1 \rightarrow A_1, A_1, A_2$. The similar explanation for $\lambda = 2$, we complete the all 1st-order and 2nd-order TV-FRGs and shown in column 2 and column 3 of **Table 4**, respectively.

Table 4 The complete the 1st-order and 2nd-order TV-FRGs.

No	1st-order TV-FRGs	2nd-order TV-FRGs
1	$A_1 \rightarrow A_1$	$(A_1, A_1) \rightarrow A_1$
2	$A_1 \rightarrow A_1, A_1$	$(A_1, A_1) \rightarrow A_1, A_2$
3	$A_1 \rightarrow A_1, A_1, A_2$	$(A_1, A_2) \rightarrow A_3$
--	---	---
20	$A_7 \rightarrow A_7, A_7$	$(A_7, A_7) \rightarrow A_7, A_7$
21	$A_7 \rightarrow A_7, A_7, A_7$	$(A_7, A_7) \rightarrow \#$
22	$A_7 \rightarrow \#$	

Step 7: Defuzzify and calculate the forecasting output values

The last step of the proposed model is to defuzzify the forecasting state to a crisp output value from fuzzy forecasting rules. In particular, to defuzzify the fuzzified data values, the our defuzzified principle in article [30] is presented to compute the forecasted value for all 1st-order and high-order time variant FRGs in training phase. Next, we use a defuzzified principle [14] for computing with the unknown linguistic value in testing phase. The forecasting principles is presented as follows:

Principle 1: Apply to calculate forecasting output value in the training phase

To calculate forecasting value output based on information of the each group. We divide each corresponding interval with respect to the fuzzy sets in the next state of the TV-FRGs into 3 sub-intervals with the same length. The forecasted value for each group is defined as follows:

$$\text{Forecasted\_value} = \frac{1}{2*n}\sum_{i=1}^{n}(subm_{ik} + Value\_lu_{ik}) \tag{9}$$

Where, n denotes the total number of fuzzy sets on the next state of TV-FRG.

- ✓ $subm_{ik}$ denotes the midpoint value of one of 3 sub-intervals ($1 \le k \le 3$) with respect to i-th linguistic value in the next state of FRG that the real data at forecasting time falls into this sub-interval.
- ✓ $Value\_lu_{ik}$ is one of 2 values belongs to the lower bound and upper bound value of one of 3 sub-intervals which has the real data at forecasting time falls within sub-interval u_{ik} (i.e., $u_{ik} = [L_{ik}, U_{ik}]$).
  - **If** the real data value at forecasting time minors the midpoint value of sub-interval u_{ik}, then $Value\_lu_{ik}$ is assigned as the lower bound of sub-interval u_{ik}; **else** $Value\_lu_{ik}$ is assigned as the upper bound of sub-interval u_{ik}.

For example, suppose that we want to calculate the forecasting value in year 1972. Based on column 4 of **Table 5** shown that the 2nd-order fuzzy relationship group G1 ($A_1 \rightarrow A_1$) is formed from a FLR with next state A_1 appearing at year 1972, where the highest membership degree of A_1 fall into interval $u_1 = [13000, 14029.68)$. Thus, we partition the interval u_1 into 3 sub-intervals which are $u_{1.1} = [13000, 13343.23)$, $u_{1.2} = [13343.23, 13686.45)$ and $u_{1.3} = [13686.45, 14029.68)$, respectively. In addition, the historical data of year 1972 with respect to linguistic value A_1 of 13563 and it fall within sub-interval $u_{1,2} = [13343.23, 13686.45)$. The The corresponding mid-value for the sub-interval $u_{1.2}$ is $subm_{1.2}$ (13514.84). Following, the value of $Value\_lu_{ik}$ obtained by comparing between the real value of year 1972 and the midpoint value of sub-interval $u_{1.2}$. By this way, the value of Val_lu_{ik} (Val_lu_{21}) is assigned equal to 13686.45. Finally, the forecasting output value of year 1972 can be computed according to Eq. (10) as follows:

$$\text{Forecasted \_value} = \frac{1}{2}(13514.84 + 13686.45) = 13600.65 \tag{10}$$

By the same way, we can get the forecasting value for the 2nd-order fuzzy relationship group ((A_1, A_1) → A_1) appearing at year 1973 as 13943.9.

Principle 2: Using for calculating forecasting output value in the testing phase

For testing phase, we calculate forecasted value for a group of fuzzy relationship which has the unidentified linguistic value on the right-hand side based on the master vote scheme [14]. Assume there a λ-order fuzzy relationship group as $A_{t-\lambda}$, $A_{t-(\lambda+1)}$, A_{t1} → #, the forecasting value is estimated according to Eq. (11), where the symbol w_h is the highest votes predefined by user for each other problem, λ is the order of the FLRs, the symbols $M_{t-1}, M_{t-2}\ldots$ and $M_{t-\lambda}$ are the middle values of the corresponding intervals which related to the latest fuzzy set and other fuzzy sets on the left-hand side of fuzzy relationship group having the maximum membership values of $A_{t-1}, A_{t-2}, \ldots$, and $A_{t-\lambda}$ occur at intervals $u_{t1}, u_{t2}, \ldots$, and $u_{t-\lambda}$, respectively.

$$Forecated\text{d\_value} = \frac{(M_{t-1} * w_h) + M_{t-2} + \cdots + M_{t-\lambda}}{w_h + (\lambda - 1)} \tag{11}$$

Based on the 2 forecasting principles above and fuzzy relationship groups in **Table 4**, we complete forecasting results for the enrolments the period from 1971 to 1992 based on 1st-order and high-order time variant FRGs under 7 intervals are shown in column 4 and 5 of **Table 5**, respectively.

Table 5 The complete forecasted outputs based on the 1st-order and 2nd-order FTS.

Year	Actual data	Fuzzy set	1st-order forecasted value	2nd-order forecasted value
1971	13055	A1	---	---
1972	13563	A1	13600.6	---
1973	13867	A1	13772.3	13943.9
1974	14696	A2	14095.7	14343.2
---	---	---	---	---
1991	19337	A7	19308.4	19308.4
1992	18876	A7	19124	19031.8
MSE			**129623.34**	**70188.37**

The efficiency of the proposed forecasting model is evaluated using various statistical indexes, namely Mean Square Error (MSE) and Mean Absolute Percentage Error (RMSE). The lower value of MSE, RMSE indicates better performance of proposed model. The evaluation criterions are determined by the following equations:

$$\mathrm{MSE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i}=\lambda}^{\mathrm{n}}(\mathrm{F_i}-\mathrm{R_i})^2 \tag{12}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=\lambda}^{\mathrm{n}}(\mathrm{F_i}-\mathrm{R_i})^2} \tag{13}$$

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

Stage 3: A hybrid FTS forecasting model based on combining the HA and PSO

In this section, we present the hybrid FTS forecasting model by combining HA and PSO algorithm with the aim to improve forecasting accuracy. In which, PSO algorithm is used to minimize the MSE value by adjusting the lengths of the initial intervals which are determined by HA and membership values, respectively. The proposed forecasting model is named FTSHA-PSO. The briefly explanation of the FTSHA-PSO are given as following. In the FTSHA-PSO model, for the training phase, each particle is used to represent the partitioning of time series data (e.g., n intervals). Assume that the lower bound and upper bound of the universe of discourse U be x_0 and x_n, respectively. Each particle denotes a vector containing of n-1 elements as x_1, x_2, …, x_{n-2} and x_{n-1}, where $(1 \leq i \leq n-1)$ and $x_i \leq x_{i+1}$. From these n-1 elements, define the n intervals as $u_1 = [x_0, x_1]$, $u_2 = [x_1, x_2]$,,..., $u_i = [x_{i-1}, x_i]$,… and $u_n = [x_{n-1}, x_n]$, respectively. In case of movement of particle in a swarm from 1 position to another position, the elements of the corresponding new array always require to be adjusted in an ascending order such that $x_1 \leq x_2 \leq \cdots \leq x_{n-1}$. In processing for the training phase, the FTSHA-PSO model permits each particle moving form current position to other position by Eqs. (5) and (6), and repeats the steps until the stopping criterion is satisfied. If the stopping criterion is satisfied, then all the FRGs obtained by the global best position (G_{best}) among all personal best positions (P_{best}) of all particles which used to forecast the new testing data in testing phase. Here, the MSE value in Eq. (12) is used to evaluate the forecasted accuracy of each particle. The complete steps of the proposed model are presented in Algorithm 2 as follows:

Algorithm 2 The FTSHA-PSO algorithm in the training phase

1 **Input**: Historical time series data
2 **Output**: The forecasting results and the MSE value (MSE = Gbest = min(Pbest))
Begin
3 Define the initial intervals by applying HA and **use** forecasting steps in Subsection above to reach the initial forecasting accuracy (MSE).
4. Initialize: a population of N particles
- ✓ The initial position X_{ki} and the velocity V_{ki} of all particles, respectively.
- ✓ The initial personal best position vectors of the ^{k}th particle is the same as its initial position vector at the beginning and find Gbest

5. **do**
5.1**. for** each particle k, ($1 \leq k \leq N$) **do**
- ✓ Following the steps in the part above sequentially, from step 3 to step 7 such as: defining linguistic terms, fuzzify all historical, determining all $\lambda -$ order fuzzy logical relationships, establishing all $\lambda -$ order TV-FRGs, defuzzify forecasting values, calculating the MSE value for each particle
- ✓ The new Pbest of all particles is saved according to the MSE values.

end for
5.2. The new Gbest of all particles is saved according to the MSE values
6. **for** each particle k, ($1 \leq k \leq N$) **do**
- ✓ The particle k is moved to another position according to Eqs. (5) and (6)

end for
- ✓ Update ω according to Eq. (4)

while (the maximum moving steps(*iter_max*) or the minimum MSE are reached)
End.

Algorithm 3 The FTSHA-PSO algorithm in the testing phase

The proper lengths of intervals and order of FLRs obtained in Algorithm 2 that are used to estimate untrained data in the testing phase based on the Principle 2 in the proposed model.

Example: The illustrating of the FTSHA-PSO model in the training phase is presented as follows. In this example, let the number of intervals and particles be 7 and 4 respectively, and the FTSHA-PSO model uses the PSO to obtain all the 2nd-order FLRs by adjusting the length of intervals for the historical enrolments data. In proposed FTS model, we have the universe of discourse on U = [13000, 20000], where lower bound x_0 = 13000 and upper bound x_7 = 20000, respectively. For finding the optimal solution, we define values for the parameters used in Eqs. (5) and (6) as: The range of x_{ki} be limited to (13000, 20000], the range of v_{ki} be limited to [–100, 100], the values of C_1 and C_2 be 2, and the value of ω = 0.9 (where ω linearly decreases its value to the lower bound, 0.4, through the whole training process) and maximum number of iterations be 2, respectively. The positions and velocities of all particles are initialized randomly and listed in **Tables 6** and **7**, respectively.

In **Table 6**, we have shown the 7 intervals for each particle which are $u_1 = [x_0, x_1]$, $u_2 = [x_1, x_2]$,..., $u_i = [x_{i-1}, x_i]$,… and $u_n = [x_{n-1}, x_n]$, respectively. Where, the intervals of the initial position of particle 1 are established as the same the one which are created from HA in Subsection 3.1 and listed as u_1 = [13000, 14029.68), u_2 = [14029.68, 14980.2), u_3 = [14980.2, 15857.5), u_4 = [15857.5, 16808), u_5 = [16808, 17605), u_6 = [17605, 18340.16), u_7 = [18340.16, 20000]. Next, we follow the steps of Algorithm 2 and achieve the optimal intervals which are utilized for obtaining the forecasting results. The MSE value of particle1 is calculated based on Eq. (12). The MSE values for the remaining 3 particles are found in a similar manner. Based on the corresponding MSE value, every particle updates its own personal best positions (Pbest) so far. For simplicity, the initial Pbests are considered for the initial positions of all particles. The Pbests of all particles so far are shown in **Table 8**. From **Table 8**, the global best position Gbest = min(Pbest) is created by particle 3, because its MSE value is the least among all particles so far. After the 1st iteration, all particles move to the 2nd positions according to Eqs. (5) and (6). The 2nd positions and the corresponding new MSE values of all particles are presented in **Table 9**.

Table 6 Randomly generated initial positions of all particles.

Particle	x_1	x_2	x_3	x_4	x_5	x_6	MSE
1	**14029.68**	**14980.2**	**15857.5**	**16808**	**17605**	**18340.16**	**70185.73**
2	13485.34	14156.19	14217.29	18109.05	18305.1	19046.01	426939.77
3	13572.29	14395.55	15206.81	15572.54	16668.41	17504.68	31861.41
4	14368.55	15098.79	15672.91	16495.91	17598.12	18025.22	55020.22

Table 7 Randomly generated initial velocities of all particles.

Particle	v_1	v_2	v_3	v_4	v_5	v_6
1	–4.45	–75.51	–40.73	–20.91	39.05	–93.04
2	13.48	–92.09	76.32	86.76	52.71	–58.5
3	–82.46	14.13	–45.76	55.36	46.77	–38.01
4	–31.41	–60.12	–26.92	–75.18	77.02	3.83

Table 8 The initial Pbest of all particles; the global best position is created by particle 3 as its MSE is the least among 4 particles.

Particle	x_1	x_2	x_3	x_4	x_5	x_6	MSE
1	14029.68	14980.2	15857.5	16808	17605	18340.16	70185.73
2	13485.34	14156.19	14217.29	18109.05	18305.1	19046.01	426939.77
3	13572.29	14395.55	15206.81	15572.54	16668.41	17504.68	31861.41
4	14368.55	15098.79	15672.91	16495.91	17598.12	18025.22	55020.22

Table 9 The 2nd positions of all particles.

Particle	x_1	x_2	x_3	x_4	x_5	x_6	MSE
1	13929.68	14880.2	15757.5	16708	17505	18240.16	54313.57
2	13565.06	14256.19	14317.29	18009.05	18205.1	18946.01	402776.58
3	13498.08	14408.27	15165.63	15622.36	16710.5	17470.47	36398.23
4	14268.55	14998.79	15572.91	16395.91	17498.12	17925.22	70457.58

By comparing the MSE values shown in **Table 8** with those listed in **Table 9**, it is obvious that particle 1 and particle 2 in **Table 9** attained a better position than their own Pbest values so far. Thus, the 2 particles update their own Pbest values, which are shown in **Table 10**. The new Gbest is obtained by particle 3, because its MSE value is the least among all the particles so far. The FTSHA-PSO model is accomplished by repeating the above steps until the maximum number of iterations is reached. Finally, the proper lengths of intervals are achieved by the Gbest value that the particle 3 attains so far, and are employed for obtaining the final forecasting. These obtained results used to forecast the new testing data in the testing phase based on Algorithm 3.

Table 10 The 2nd Pbest of all particles and the Gbest value is obtained by particle 3.

Particle	x_1	x_2	x_3	x_4	x_5	x_6	MSE
1	13929.68	14880.2	15757.5	16708	17505	18240.16	54313.57
2	13565.06	14256.19	14317.29	18009.05	18205.1	18946.01	402776.58
3	13572.29	14395.55	15206.81	15572.54	16668.41	17504.68	**31861.41**
4	14368.55	15098.79	15672.91	16495.91	17598.12	18025.22	55020.22

Results and discussion

In this paper, the proposed model is applied to forecast enrollments of University of Alabama [4]. The forecasting results of proposed model are compared with those of corresponding models in the literature for various order and different intervals. For implementing the experiment, we use visual studio 2019 environment with C# programming language on an Intel Core i7 PC with 8GB RAM. Based on parameters in **Table 11**, the FTSHA-PSO model is executed 20 times on enrolments dataset with various number of orders and intervals. Then, the best result of all runs is recorded to be the final forecast result. All forecasted accuracies are evaluated by MSE and RMSE value which are presented in Eqs. (12) and (13), respectively.

Table 11 The parameters of PSO used in the FTSHA-PSO model for forecasting enrolments.

The parameters in PSO	Values of parameters
The number of particles N =	50
The max number of iterations iter_max =	200
The value of inertial weight ω is decreased by	ω_{max} = 0.9 to ω_{min} = 0.4
The coefficient C1 = C2 =	2
The position in search range: **X =**	[13000, 20000]
The velocity in search range: **V =**	[–100, 100]

Forecasting results based on the 1st-order fuzzy time series

In order to evaluate the effectiveness of the proposed model based on the 1st-order FTS with the number of intervals equal to 7, the forecasting models is introduced in articles [33-36,27,29] were considered for comparison. From the parameters are set for the enrolments data. A comparison in term of RMSE value between the FTSHA-PSO model and its counterparts are shown in **Table 12**. Based on forecasting results in **Table 12**, the FTSHA-PSO model gets the smallest RMSE value of **172.9** among all the compared models. Differences between the FTSHA-PSO model and models mentioned above is the way which the fuzzy relationship group and method of partitioning the universe of discourse are applied to establish the forecasting model. Three forecasting models in works [33-35] are constructed based on Chen's model to forecast different problems and apply information granules to partition, while the FTSHA-PSO model uses HA for determining unequal-sized interval lengths. In addition, 2 models in articles [27,29] based on HA and Chen's fuzzy relationship groups to structure the forecasting model, while the FTSHA-PSO model uses an approach that benefits from the concept of time-variant FRG [17] to establish the forecasting model. Finally, the FTSHA-PSO is different from the model [36] in the partitioning the universe of discourse they used, the former uses the HA which combines with PSO for finding the optimal interval lengths but latter utilizes the maximum spanning tree based fuzzy clustering algorithm for partitioning intervals with different lengths in the intuitionistic FTS forecasting model.

Table 12 A comparison of the forecasting results of the FTSHA-PSO model with its counterparts based on 1st-order FTS under 7 intervals.

Year	Real data	[33]	[34]	[35]	[27]	[36]	[29]	FTSHA-PSO
1972	13563	13486	13944	14279	13820	13500	13865	13619.24
1973	13867	14156	13944	14279	13820	14155	14082	13729.16
---	---	---	---	---	---	---	---	---
1991	19337	18808	18933	19257	19135	19575	19165	19321.66
1992	18876	18808	18933	19257	19135	18855	15219	19167.49
RMSE		**578.3**	**506**	**445.2**	**441.3**	**350.9**	**210.9**	**172.9**

In addition, the FTSHA-PSO model is also executed 20 runs to be compared with various 1st-order FTS models under number of intervals of 14 intervals. Five forecasting models are presented in research works [4,6,11,14,15,17,35] were selected for comparison. A comparison of the forecasted results is shown in **Table 13** where the number of intervals is 14 for all forecasting models. At the same intervals, it is obvious that the FTSHA-PSO model has the MSE value 5123 which is the lowest among all forecasting models compared.

Table 13 A comparison of the forecasting results of the FTSHA-PSO model with its counterparts based on 1st-order FTS under number of intervals of 14.

Year	real data	[4]	[6]	[11]	[14]	[15]	[17]	[35]	FTSHA-PSO
1971	13055								
1972	13563	14000	13653	13714	13555	13579	13434	13512	13558.12
1973	13867	14000	13653	13714	13994	13812	13841	13998	13863.28
----	----	----	----	----	----	----	----	-----	----
1991	19337	19000	19059	19149	19340	19260	19340	19666	19262.03
1992	18876	19000	19059	19014	19014	19031	18820	18718	19035.97
MSE		**407507**	**31684**	**35324**	**22965**	**8224**	**7475**	**14534**	**6665.89**
RMSE		**638.4**	**178**	**187.9**	**151.5**	**90.7**	**86.5**	**120.6**	**81.6**

Forecasting results based on the high-order fuzzy time series

In this subsection, all historical enrolments dataset [4] covering a period from year 1971 to 1992 are partitioned into 2 parts to implement comparison results of the FTSHA-PSO with the ones of the existing methods, based on various high-orders. The 1st part including 19 observations from year 1971 to 1989 is used as training data set and the 2nd part consists of 3 observations is used as the testing data set. The performance of the FTSHA-PSO and the compared models are evaluated using the MSE and RMSE function.

Experimental results in the training phase

The FTSHA-PSO model is evaluated through the different high-order FLRs of fuzzy time series. In particular, in order to verify the superiority in the forecasted accuracy of the FTSHA-PSO model with number of intervals equal to 7, the accuracies from cited papers in [7,12,14,15,17] are selected for comparing. A comparison of the forecasting results is listed in **Table 14**, in what the number of intervals is established for all forecasting models equal to 7. At the same intervals 7, the FTSHA-PSO model gets the lowest MSE values which are 12457.8, 529.54, 443.47, 412.39, 366.42, 286.26, 163.27 and 371.13 for 2nd-order, 3rd-order, 4th-order, 5th-order, 6th-order, 7th-order, 8th-order and 9th-order fuzzy time series, respectively. It can be seen that the FTSHA-PSO model achieves more precise than any other existing models under different high-order fuzzy relationships at all. Among all fuzzy relationships is done in the model, the FTSHA-PSO model obtains the lowest MSE value of 14420.4 with 8th-order fuzzy relationships. The major difference between FTSHA-PSO model and the compared models is in establishing fuzzy relationship groups and optimization technique they used. In optimization method, the model [12] performs genetic algorithm but the models in articles [14,15,17] and the FTSHA-PSO model proceed the PSO algorithm to achieve the best intervals, respectively. In addition to using PSO to find suitable intervals, the FTSHA-PSO model incorporates HA to partition the different initial intervals of the Universe of discourse instead of equal length intervals. In the determining of fuzzy relationship groups, the FTSHA-PSO model is constructed from model [17], the remaining models in articles [7,12,14,15] are designed based on Chen's structure [4]. From the above analysis, it can be seen that the FTSHA-PSO model gives more convincing forecasting results compared to its counterparts based on the high-order FTS.

Table 14 A comparison of the results obtained between the FTSHA-PSO model and its counterparts based on the various high-order FTS with 7 intervals.

Order	[7]	[12]	[14]	[15]	[17]	FTSHA-PSO
2	89093	67834	67123	19594	19868	12457.8
3	86694	31123	31644	31189	31307	529.54
4	89376	32009	23271	20155	23288	443.47
5	94539	24984	23534	20366	23552	412.39
6	98215	26980	23671	22276	23684	366.42
7	104056	26969	20651	18482	20669	286.26
8	102179	22387	17106	14778	17116	**163.27**
9	102789	18734	17971	15251	17987	371.13
Average MSE	**95867.63**	**31377.5**	**28121.38**	**20261.38**	**22183**	**1878.79**

In addition, the FTSHA-PSO model is also compared with its counterparts which are introduced in papers [7,12,14,15,37] based on the different high-order fuzzy time series with number of intervals of 14. From the parameters are expressed in **Table 6**. The FTSHA-PSO model is executed 20 runs, and the best result of runs is taken to be the final result. A comparison of the forecasting accuracy with various high-orders and the same number of intervals between the FTSHA-PSO model and its counterparts are listed in **Table 15**. Where, the FTSHA-PSO model is different from the model [37] in the way that the method of constructing forecasted model they used. The former applies the concept of the time-variant fuzzy relationship group but the latter proceeds the adaptive time-variant fuzzy time series to establish the forecasting model, respectively. In addition, the forecasting model in article [7] and the FTSHA-PSO model, both of them use the 5^{th}-order fuzzy relationship but our FTSHA-PSO model is much more superior in term of forecasting accuracy. Remaining forecasting models in articles [12,14,15], they use the fuzzy logical relationship with number of orders is larger, but the results obtained from our model are also better than the existing competing models. In particular, from **Table 15**, it is obvious that our forecasting model gets the MSE value of **18** which is the smallest among all compared forecasting models. This can conclude that the proposed model provides the superior forecasting performance compared to its counterparts based on the various high-order FLRs at all.

Table 15 A comparison of the forecasting results obtained between the FTSHA-PSO model and its counterparts based on the various high-order FTS with 14 intervals.

Years	Real data	[7]	[37]	[12]	[14]	[15]	FTSHA-PSO
1971	13055						
1972	13563						
1973	13867		14934.5				
1974	14696		15590				
1975	15460		15422.9				
1976	15311	15500	15603				15314
1977	15603	15500	15861				15608
1978	15861	15500	16807				15858
1979	16807	16500	16919	16846			16803
1980	16919	16500	16388	16846	16890	16920	16920
1981	16388	16500	15553.9	16420	16395	16388	16390
---	---	---	---	---	---	---	---
1991	19337	19500	18876	19334	19337	19335	19332
1992	18876	18500	14934.5	18910	18882	18882	18876
MSE		**86694**	**53084**	**1101**	**234**	**173**	**18**
RMSE		**294.44**	**230.4**	**33.18**	**15.3**	**13.15**	**4.24**

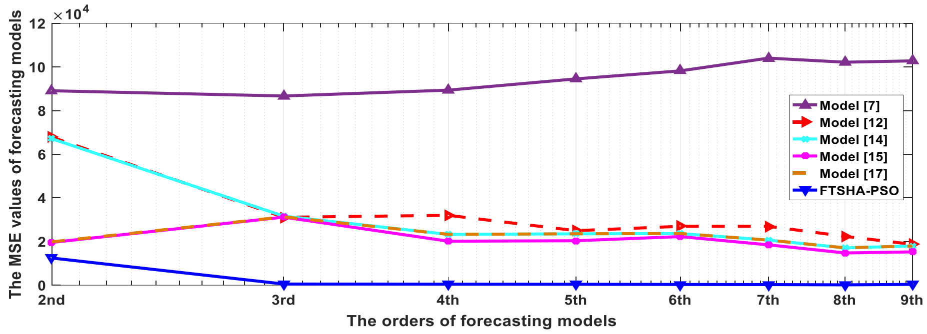

Figure 4 The curves present the MSE values between the FTSHA-PSO model and its counterparts based on the various high-order FTS.

To be easily visualized, **Figure 4** depicts the trend in term of forecasting accuracy between the FTSHA-PSO model and its counterparts with different high-order FLRs. From these curves, it can be seen that forecasting accuracy of the FTSHA-PSO model is more precise than those of compared models under different high-order FLRs at all. To sum up, demonstrations above show that the FTSHA-PSO model outperform the existing models based on high-order FTS model with different number of intervals in forecasting the enrolments of University of Alabama.

Experimental results in the testing phase

Based on the historical enrolments data for the past years, we can forecast the new enrollment for the next year only. For example, the historical data of enrollments from year 1971 to 1989, is utilized to forecast the new enrollment of year 1990. Similarly, a new enrolment of year 1991 can be forecasted based on the enrollments from years 1971 - 1990. After the training data have been well trained by the FTSHA-PSO, the future enrollment values could be accomplished to compare with the ones of the forecasting models presented in articles [4,11,14,37]. The comparison of the forecasted results produced based on the 3rd-order FLR the number of different intervals and the highest vote W_h = 20 (constant-defined by the user) in **Tables 16** and **17**. From **Tables 16** and **17**, it can be seen that the FTSHA-PSO model obtains the lowest RMSEs value of 98.6 and 72.53 among 5 compared models, respectively. From these results indicate that the our FTSHA-PSO model is more precise than its counterparts based on 3rd-order FTS with different number of intervals.

Table 16 A comparison of the forecasting results between the FTSHA-PSO model and other models with the number of intervals = 7 and which use vote W_h = 20.

Year	**Real data**	**[4]**	**[11]**	**[14]**	**ATVF-KM [37]**	**ATVF-PSO [37]**	**FTSHA-PSO**
1990	19328	18168	18059	18326	19525	19226	19308
1991	19337	18909	18669	19212	19150	19182	19351.9
1992	18876	19609	19083	19203	18933	18876	19045
RMSE		**773.66**	**576.66**	**484.16**	**160.43**	**107.29**	**98.6**

Table 17 A comparison of the forecasting results between the FTSHA-PSO model and other models with the number of intervals = 14 and which use vote W_h = 20.

Year	**Real data**	**[4]**	**[11]**	**[14]**	**ATVF-KM [37]**	**ATVF-PSO [37]**	**FTSHA-PSO**
1990	19328	18162	17862	18120	19287	19238	19230
1991	19337	18721	18633	19027	18811	19224	19336.74
1992	18876	19221	19085	19137	18836	19224	18954.6
RMSE		**709**	**653.66**	**621.91**	**305.7**	**92.35**	**72.53**

Conclusions

In this paper, we propose a hybrid FTS forecasting model combining HA and PSO, namely FTSHA-PSO. The FTSHA-PSO model has addressed 2 issues are considered to be important and greatly affect the forecasting accuracy that is the problem about determining of length of intervals and how to establish fuzzy relationship groups. To overcome the limitations of fuzzy time series models using the fuzzy relationship groups, the FTSHA-PSO model uses the concept of time variant fuzzy relationship group to calculate the forecasting results output. Using this fuzzy relationship group has proved to be more appropriate for practical use. In addition, the PSO optimization technique is applied in finding the optimal lengths of intervals from the universe of discourse to improve the forecast accuracy of the FTSHA-PSO model. Among mining techniques and finding the best solution, PSO is considered to perform better compared to other heuristic techniques in terms of success rate and solution quality. By combining HA and PSO technique, the forecasting efficiency of the FTSHA-PSO model can be significantly improved. The experimental results on dataset of University of Alabama show that, in many cases, the FTSHA-PSO model gets better forecasting performance than the existing ones. Details of the comparison are shown in **Tables 12 - 17**. Although our FTSHA-PSO model shows that the superior forecasting capability compared with existing forecasting models based on the high-order FLRs. However, determining FLR in high-order is more complicated and computationally more expensive than 1^{st}-order. Therefore, development of new approaches that can automatically find the optimal order of the high-order FLRs is a worthy idea in FTS forecasting model. Those will be the future work closely related to this research.

Acknowledgment

This work was supported by Thai Nguyen University of Technology (TNUT).

References

[1] LA Zadeh. Fuzzy sets. *Inf. Control* 1965, **8**, 338-53.
[2] Q Song and BS Chissom. Fuzzy time series and its models. *Fuzzy Sets Syst*. 1993; **54**, 269-77.
[3] Q Song and BS Chissom. Forecasting enrollments with fuzzy time series - Part I. *Fuzzy Sets Syst*. 1993; **54**, 1-9.
[4] SM Chen. Forecasting enrollments based on fuzzy time series. *Fuzzy Sets Syst*. 1996; **81**, 311-9.
[5] HK Yu. Weighted fuzzy time series models for TAIEX forecasting. *Physica A* 2005; **349**, 609-24.
[6] VR Uslu, E Bas, U Yolcu and E Egrioglu. A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relations. *Swarm Evol. Comput*. 2014; **15**, 19-26.
[7] SM Chen. Forecasting enrollments based on hight-order fuzzy time series. *Cybern. Syst*. 2002; **33**, 1-16.
[8] K Huarng. Effective lengths of intervals to improve forecasting in fuzzy time series. *Fuzzy Sets Syst*. 2001; **123**, 387-94.
[9] LW Lee, LH Wang, SM Chen and YH Leu. Handling forecasting problems based on two-factors high-order fuzzy time series. *IEEE Trans. Fuzzy Syst*. 2006; **14**, 468-77.
[10] SM Chen and K Tanuwijaya. Fuzzy forecasting based on high- order fuzzy logical relationships and automatic clustering techniques. *Expert Syst. Appl*. 2011; **38**, 15425-37.
[11] SM Chen and NY Chung. Forecasting enrollments of students by using fuzzy time series and genetic algorithms. *Int. J. Inf. Manag. Sci*. 2006; **17**, 1-17.
[12] SM Chen and NY Chung. Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *Int. Intell. Syst*. 2006; **21**, 485-501.
[13] LW Lee, LH Wang and SM Chen. Temperature prediction and TAIFEX forecasting based on hight order fuzzy logical ralationship and genetic simulated annealing techniques. *Expert Syst. Appl*. 2008; **34**, 328-36.
[14] IH Kuo, SJ Horng, TW Kao, TL Lin, CL Lee and Y Pan. An improved method for forecasting enrollments based on fuzzy time series and particle swarm optimization. *Expert Syst. Appl*. 2009; **36**, 6108-17.
[15] YL Huang, SJ Horng, M He, P Fan, TW Kao, MK Khan, JL Lai and IH Kuo. A hybrid forecasting model for enrollments based on aggregated fuzzy time series and particle swarm optimization. *Expert Syst. Appl*. 2001; **38**, 8014-23.
[16] LY Hsu, SJ Horng, TW Kao, YH Chen, RS Run, RJ Chen, JL Lai and IH Kuo. Temperature prediction and TAIFEX forecasting based on fuzzy relationships and MTPSO techniques. *Expert Syst. Appl*. 2010; **37**, 2756-70.

[17] NC Dieu and NV Tinh. Fuzzy time series forecasting based on time-depending fuzzy relationship groups and particle swarm optimization. *In*: Proceedings of the 9th National Conference on Fundamental and Applied Information Technology Research, Can Tho, Vietnam. 2016, p. 125-33.
[18] JI Park, DJ Lee, CK Song and MG Chun. TAIFEX and KOSPI 200 forecasting based on two-factors high-order fuzzy time series and particle swarm optimization. *Expert Syst. Appl*. 2010; **37**, 959-67.
[19] SM Chen and BHP Dang. Fuzzy time series forecasting based on optimal partitions of intervals and optimal weighting vectors. *Knowl. Based Syst*. 2016; **118**, 204-16.
[20] SM Chen and WS Jian. Fuzzy forecasting based on two-factors second-order fuzzy-trend logical relationship groups, similarity measures and PSO techniques. *Inf. Sci*. 2017; **391**-**392**, 65-79.
[21] M Bose and K Mali. A novel data partitioning and rule selection technique for modelling high-order fuzzy time series. *Appl. Soft Comput*. 2017; **63**, 87-96.
[22] Z Tian, P Wang and T He. Fuzzy time series based on K-means and particle swarm optimization algorithm. *In*: Proceedings of the International Conference on Man-Machine-Environment System Engineering, Xi'an, China. 2016, p. 181-9.
[23] Z Zhang and Q Zhu. Fuzzy time series forecasting based on k-means clustering. *Open J. Appl. Sci*. 2012; **2**, 100-3 .
[24] NV Tinh and NC Dieu. Improving the forecasted accuracy of model based on fuzzy time series and k-means clustering. *J. Sci. Technol.* 2017; **3**, 46-55.
[25] E Bulut, O Duru and S Yoshida. A fuzzy time series forecasting model formulti-variate forecasting analysis with fuzzy c-means clustering. *World Acad. Sci. Eng. Technol*. 2012; **63**, 765-71.
[26] S Askari and N Montazerin. A high-order multi-variable Fuzzy Time Series forecasting algorithm based on fuzzy clustering. *Expert Syst. Appl*. 2015; **42**, 2121-35.
[27] NC Ho, NC Dieu and VN Lan. The application of hedge algebras in fuzzy time series forecasting. *J. Sci. Technol*. 2016; **54**, 161-77.
[28] NC Ho and W Wechler. Hedge algebra: An algebraic approach to structures of sets of linguistic truth values. *Fuzzy Sets Syst*. 1990; **35**, 281-93.
[29] H Tung, ND Thuan and VM Loc. The partitioning method based on hedge algebras for fuzzy time series forecasting. *J. Sci. Technol*. 2016; **54**, 571-83.
[30] NV Tinh. Enhanced forecasting accuracy of fuzzy time series model based on combined fuzzy C-mean clustering with particle swam optimization. *Int. J. Comput. Intell. Appl*. 2020; **19**, 2050017.
[31] J Kennedy and R Eberhart. Particle swarm optimization. *In*: Proceedings of the IEEE International Conference on Neural Network, Perth, WA, Australia. 1995, p. 1942-8.
[32] NC Ho, TT Son and DT Long. Hedge algebras with limited number of hedges and applied to fuzzy classification problems. *Vietnam J. Sci. Technol*. 2012; **48**, 13-22.
[33] L Wang, XD Liu and W Pedrycz. Effective intervals determined by information granules to improve forecasting in fuzzy time series. *Expert Syst. Appl*. 2013; **40**, 5673-79.
[34] L Wang, X Liu, W Pedrycz and Y Shao. Determination of temporal information granules to improve forecasting in fuzzy time series. *Expert Syst. Appl*. 2014; **41**, 3134-42.
[35] W Lu, X Chen, W Pedrycz, X Liu and J Yang. Using interval information granules to improve forecasting in fuzzy time series. *Int. J. Approx. Reason*. 2015; **57**, 1-18.
[36] Y Wang, Y Lei, X Fan and Y Wang. Intuitionistic fuzzy time series forecasting model based on intuitionistic fuzzy reasoning. *Math. Probl. Eng*. 2016; **2016**, 5035160.
[37] K Khiabani and SR Aghabozorgi. Adaptive time-variant model optimization for fuzzy-time-series forecasting. *IAENG Int. J. Comput. Sci*. 2015; **42**, 107-16.