American Journal of Information Science and Computer Engineering
Vol. 5, No. 3, 2019, pp. 94-102
http://www.aiscience.org/journal/ajisce
ISSN: 2381-7488 (Print); ISSN: 2381-7496 (Online)

An Improved Fuzzy Time Series Forecasting Model Based on Combining K–means Clustering with Harmony Search

Nghiem Van Tinh*

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

Abstract

Various methods have been presented to investigate the length of data interval and partition number of universe of discourse in fuzzy time series (FTS) model to achieve high level forecasting accuracy. These resolve the fluctuations of FTS as mentioned by previous researchers. However, it still has difficulties in choosing the optimal length of interval. In this paper, we present a hybrid forecasting model based on combing K-mean clustering and Harmony search algorithm (HAS) to overcome the difficulties mentioned above. Firstly, the K-mean clustering algorithm is used to divide the historical data into clusters and adjust them into intervals with different initial length, then the strong global searching ability of particle swarm optimization is also used to adjust initial length and obtain the optimal length of interval universe of discourse with the aim to increase forecasting accuracy of model. Finally, two numerical datasets (enrollments data of the University of Alabama, and yearly deaths in car road accidents in Belgium) are used to verify the feasibility of the model by comparing and analyzing the forecasting accuracy between proposed model and other forecasting methods. The empirical analysis not only demonstrates the forecasting procedure and the way to obtain the suitable length of interval, but also shows that the proposed model significantly outperforms the conventional counterparts.

Keywords

Forecasting, Fuzzy Time Series, K–mean Clustering, Harmony Search, Enrolments

Received: July 9, 2019 / Accepted: August 28, 2019 / Published online: September 6, 2019

1. Introduction

Advance forecasting of future events of time series has always influence people during the past decades. Therefore, numerous FTS models have been proposed to deal with various domain problems such as: enrollments fore-casting [1-10], crop production prediction [6, 11], stock markets prediction [5, 12-14] and temperature forecasting [14, 15]. There is the matter of fact that the traditional forecasting models such as regression analysis, moving average, autoregressive moving average and ARIMA model cannot deal with the forecasting problems in which the historical data are represented by linguistic values. Fuzzy set theory was firstly presented by Zadeh to handle problems with linguistic values. The concepts of fuzzy sets have been successfully applied to time series by Song and Chissom. They introduced both the time-invariant fuzzy time series [7] and the time-variant time series [8] model which use the max–min operations to forecast the enrolments of the University of Alabama. Unfortunately, their method needs max–min composition operations to deal with fuzzy rules. It takes a lot of computation time when fuzzy rule matrix is big. Therefore, Chen [2] proposed the first-order fuzzy time series model by using simple arithmetic calculations instead of max-min composition operations [7-9] for better forecasting accuracy. After that, fuzzy time series has been widely

* Corresponding author
E-mail address: nghiemvantinh@tnut.edu.vn

studied for improving accuracy of forecasting in many applications. Huarng [5] presented effective approaches which can properly adjust the lengths of intervals to get better forecasting accuracy. Chen [3] proposed a new forecast model based on the high-order fuzzy time series to forecast the enrollments of University of Alabama. Yu [10] presented a new model which can refine the lengths of intervals during the formulation of fuzzy relationships and hence capture the fuzzy relationships more appropriately. Both the stock index and enrollment are used as the targets in the empirical analysis. Chen & Chung [1, 4] presented the first-order and high-order fuzzy time series model to deal with forecasting problems based on genetic algorithms. Singh [6, 11] presented simplified and robust computational methods for the forecasting rules based on one and various parameters as fuzzy relationships, respectively. Lee et al. [14] presented a method for forecasting the temperature and the TAIFEX based on the high-order fuzzy logical relation groups and genetic algorithm. Recently, Particle swarm optimization technique has been successfully applied in many applications. Based on Chen's model [2], Kuo et al. [16] developed a new hybrid forecasting model which combined fuzzy time series with PSO algorithm to find the proper length of each interval. Then, they continued to present a new forecast method to solve the TAIFEX forecasting problem based on fuzzy time series and PSO algorithm [17]. Some other authors who proposed some methods for the temperature prediction and the TAIFEX forecasting, based on two-factor fuzzy logical relationships [15] and use them in which combine with PSO algorithm in fuzzy time series. In addition, some study works using other hybrid techniques can be found such as: Pritpal and Bhogeswar [18] presented a new model based on hybridization of fuzzy time series theory with artificial neural network (ANN). Matarnenh et al. [19] use feed forward artificial neural network and fuzzy logic for weather forecasting achieve better results. Article in [20] introduced a high-order multi-variable model based on FTS to deal with the problem of intervals and their lengths. Sun et al., [21] design a multivariate FTS model with multiple factors. In order to simplify the calculation and refine the evolved rules, this model integrates rough set theory into the model. Abhishekh et al. have introduced a versatile forecasting model using the concept of intuitionistic fuzzy sets based on their score function and high-order forecasting. Also, Abhishekh and Kumar [31] proposed a computational method for rice production forecasting based on high-order FTS. Further, Xian et al. [32] presented a forecasting method based on artificial fish swarm optimization algorithm. Efendi et al. [33] propose a forecasting model to predict stock market by using auto-regression method and produce more accurate forecasted value.

As already mentioned in studies above, the lengths of intervals and fuzzy relationships are two important issues considered to be serious influencing the forecasting accuracy. In this paper, a forecasting model based on combing K-means clustering and HSA is presented for forecasting the number of students in enrolments of the University of Alabama. The experimental results show that the proposed method gets better forecasting performance than the existing methods based on the first – order and the high – order FTS. The rest of this paper is organized as follows:

Section 2 introduces brief review of the basic concepts of FTS and algorithms. Section 3 presents the details of the FTS forecasting model for forecasting the enrolments of the University of Alabama based on K–mean clustering and then combines with the HSA to search the effective lengths of intervals in the universe of discourse. Section 4 compares the forecasting results of the proposed method with its counterparts from the two real life datasets. Finally, Section 5 offers some conclusions.

2. Fuzzy Time Series Concepts and Algorithms

2.1. Basic Concepts of Fuzzy Time Series

Conventional time series refer to real values, but fuzzy time series are structured by fuzzy sets. 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, respectively. The definition of FTS is briefly reviewed as follows:

Definition 1: Fuzzy time series [7, 8]

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).

Definition 2: Fuzzy logic relationship (FLR) [7, 8]

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 [3, 8]

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) [2]

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:

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

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

2.2. K-means Clustering Algorithm

K-Means clustering algorithm [29], developed three decades ago is one of the most popular clustering algorithm used in variety of domains. A priori knowledge of number of clusters are must for K-means clustering algorithm. K-Means is defined over continuous data and calculates its centers iteratively [22]. Let D = $\{d_i \mid i = 1, \ldots, n\}$ be a data set having K clusters; C = $\{c_j \mid j = 1, \ldots, K\}$ be a set of K centers and E = $\{d \mid \text{d is member of cluster k}\}$ be the set of samples that belong to the kth cluster. K-Means algorithm minimizes the following function which is defined as an objective function (1).

$$J(C) = \sum_{j=1}^{K} \sum_{d_i \in c_j}^{n} \left\| d_i - \mu_j \right\|^2 \qquad (1)$$

with μ_j be the mean of cluster c_j

The main steps of K-means algorithm are as follows:

Step 1. Select an initial partition with K clusters; repeat steps 2 and 3 until cluster membership stabilizes.

Step 2. Generate a new partition by assigning each pattern to its closest cluster center.

Step 3. Compute new cluster centers.

Step 4. Repeat Steps 2 and 3 till there is no change in cluster centers

2.3. Harmony Search Algorithm (HSA)

Using a fundamental harmony search approach, an n-dimensional real vector, called harmony, shows each solution [23]. First, a starting set of harmony vectors is randomly generated and assigned in a harmony memory (HM). The next step is a new harmony generation that considers all of the present harmonies in the HM. This goal can be achieved by deploying a harmony memory consideration rate (HMCR) aligned with a pitch adjustment rule. The procedure continues with comparing newly generated harmonies to the existing harmonies. The replacement condition is satisfied if the newly generated harmony has a better state [23]. The algorithm iterates until it satisfies a defined termination criterion condition. Compared with other heuristic optimization algorithms, it behaves with excellent effectiveness and robustness and presents lots of advantages when applied to optimization problems [24]. The five fundamental steps of the algorithm are described as follows:

Step 1. Initialize the optimization problem and algorithm parameters:

$$\text{To minimize the objective function } f(x) \qquad (2)$$

subject to $x_i \in X_i$, i= 1, 2, …N

where f (x) is the objective function; x is the set of each design variable x_i; X_i is the set of the possible range of values for each design variable; N is the number of design variables.

The HS algorithm parameters including harmony memory size (HMS), harmony memory considering rate (HMCR), pitch adjusting rate (PAR), the lower bounds (Lb) and upper bounds (Ub) for each decision variable and termination criterion should also be specified in this step.

Step 2. Initialize the Harmony Memory (HM).

HM matrix is filled with as many randomly generated solution vectors as the HMS and sorted by the values of the objective function f (x), as follows:

$$HM = \begin{bmatrix} x_1^1 & \cdots & x_N^1 \\ \vdots & \ddots & \vdots \\ x_1^{HMS} & \cdots & x_N^{HMs} \end{bmatrix}_{HMS \times N} \qquad (3)$$

Step 3. Improvise a new harmony from the HM.

A new harmony vector is generated based on three rules: The memory consideration (HMCR), pitch adjustment (PAR) and random selection (1-HMCR) are presented according to formula (4) as follows:

$$X_i^{new} \leftarrow \begin{Bmatrix} x_i(k) \in \left(x_i^1, x_i^2, \ldots, x_i^k\right) \text{ if } P_{random} = 1 - HMCR \\ x_i(k) \in \left(x_i^1, x_i^2, \ldots, x_i^{HMS}\right) \text{ if } P_{memory} = HMCR \times (1 - PAR) \\ x_1 \pm rand(0,1) \times Bw \text{ if } P_{Pitch} = HMCR \times PAR \end{Bmatrix} \qquad (4)$$

Step 4. Update the HM.

On condition that the new harmony vector showed better fitness function than the worst harmony in the HM, the new harmony is included in the HM and the existing worst harmony is excluded from the HM.

Step 5. Repeat steps 3 and 4 until the termination criterion is satisfied.

3. A Forecasting Model Based on Combining K-means Clustering with HSA

3.1. The FTS Forecasting Model Based on K-means

In this section, we presents a new method for forecasting the enrolments of University of Alabama based on time series fuzzy relationship groups and K-means clustering techniques. At first, we apply K-means clustering technique to classify the collected data into clusters and adjust these clusters into contiguous intervals for generating intervals from numerical data then, based on the interval defined, we fuzzify on the historical data, determine fuzzy relationship and establish fuzzy logical relationships and finally, we defuzzify forecasting output value based on the centroid method. The all historical data of enrolments of the University of Alabama [2] are demonstrated for forecasting process. The step-wise procedure of the proposed model is detailed as follows:

Step 1: Define the universe of discourse U and partition U into initial length intervals

In this step, apply the K-means clustering to partition the historical time series data into p clusters and sort the data in clusters in an ascending sequence. In this paper, we set number of clusters of 14 (eg. p = 14), the results are as follows:

{13055, 13563}, {13867}, {14696}, {15145, 15163}, {15460, 15311, 15433, 15497}, {15603}, {15861, 15984}, {16388}, {16807}, {16859}, {16919}, {18150}, {18970, 18876}, {19328, 19337}

Step 2: Create the cluster centers and adjust the clusters into intervals.

In this step, we generate cluster center ($Center_j$) from clusters in step 1 according to Eq. (5).

$$Center_j = \frac{\sum_{i=1}^{n} d_i}{n} \tag{5}$$

where d_i is a datum in $Cluster_j$, n denotes the number of data in $Cluster_j$ and $1 \leq j \leq p$.

Then, Adjust the clusters into intervals according to the follow rules. Assume that $Center_k$ and $Center_{k+1}$ are adjacent cluster centers, then the upper bound $Ubound_j$ of $cluster_j$ and the lower bound $Lbound_{k+1}$ of $cluster_{j+1}$ can be calculated as follows:

$$Ubound_k = \frac{Center_k + Center_{k+1}}{2} \tag{6}$$

$$Lbound_{k+1} = Cluster\_UB_k \tag{7}$$

where k = 1,..., p-1. Because there is no previous cluster before the first cluster and there is no next cluster after the last cluster, the lower bound $Lbound_1$ of the first cluster and the upper bound $Ubound_p$ of the last cluster can be calculated as follows:

$$Lbound_1 = Center_1 - (Center_1 - Cluster\_UB_1) \tag{8}$$

$$Ubound_p = Center_p + (Center_p - Cluster\_LB_p) \tag{9}$$

Based on formulas above, we get 14 intervals are listed in Table 1 as follows:

Table 1. The completed result of intervals based on K-means.

No	Intervals	Midpoint of intervals
1	(13030, 13588]	13309
2	(13588, 14434]	14011
3	(14434, 15156]	14795
---	--------	----
12	(17534, 18536]	18035
13	(18536, 19128]	18832
14	(19128, 19536]	19332

Step 3: Define the fuzzy sets for observation of enrollments

Each interval in Step 1 represents a linguistic variable of "enrollments". For 14 intervals, there are 14 linguistic variables A_i ($1 \leq i \leq 14$). For example, A_1 = {very very very very few }, A_2 = {very very very few}, A_3 = {very very few}, A_4 = {very few }, A_5 = {few }, 6 = {moderate}, A_7 = {many}, A_8 = {many many}, A_9 = {very many}, A_{10} = {too many}, A_{11} = {too many many}, A_{12} = {too many many many}, A_{13} = {too many many many many} and A_{14} = {too many many many many many}. Each linguistic variable represents a fuzzy set such that the according to (10). Each historical value is fuzzified according to its highest degree of membership. If the highest degree of belongingness of a certain historical time variable, say $F(t-1)$ occurs at fuzzy set A_i, then $F(t-1)$ is fuzzified as A_i

$$\begin{aligned} A_1 &= \frac{1}{u_1} + \frac{0.5}{u_2} + \frac{0}{u_3} + \cdots + \frac{0}{u_{14}} \\ A_2 &= \frac{0.5}{u_1} + \frac{1}{u_2} + \frac{0.5}{u_3} + \cdots + \frac{0}{u_{14}} \\ &\cdots\cdots \\ A_{14} &= \frac{0}{u_1} + \frac{0}{u_2} + \cdots + \frac{0.5}{u_{13}} + \frac{1}{u_{14}} \end{aligned} \tag{10}$$

For simplicity, the membership values of fuzzy set A_i either are 0, 0.5 or 1. These values indicate the grade of membership of u_j ($1 \leq j \leq 14$) in the fuzzy set A_i ($1 \leq i \leq 14$).

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

Step 4: Fuzzify all historical data of enrolments

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.

The completed fuzzified results enrollments listed in Table 2.

Table 2. The results of fuzzification for enrolments data.

Year	Actual data	Fuzzy set	Year	Actual data	Fuzzy set
1971	13055	A1	1982	15433	A5
1972	13563	A1	1983	15497	A5
---	---	---	---	---	---
1980	16919	A11	1991	19337	A14
1981	16388	A8	1992	18876	A13

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

Based on Definition 2. To establish the α - order fuzzy relationship, we should find out any relationship which has the $\mathrm{F}(\mathrm{t}-\alpha), \mathrm{F}(\mathrm{t}-\alpha+1), \ldots, \mathrm{F}(\mathrm{t}-1) \rightarrow F\,(t)$, where $\mathrm{F}(\mathrm{t}-\alpha), \mathrm{F}(\mathrm{t}-\alpha+1), \ldots, \mathrm{F}(\mathrm{t}-1)$ 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. So, from Table 2 and base on Definition 2, we get first – order fuzzy logical relationships are shown in Table 3. Where the fuzzy logical relationship $A_i \rightarrow A_k$ means "If the enrollment of year i is A_i, then that of year i + 1 is A_k", where A_i is called the current state of the enrollment, and A_k is called the next state of the enrollment.

Table 3. The complete results for the first- order fuzzy relationships.

No	Relationships	No	Relationships
1	A1 -> A1	11	A8 -> A5
2	A1 -> A2	12	A5 -> A5
---	--------	----	------
9	A9 -> A11	19	A13 -> A14
10	A11 -> A8	20	A14 -> A14
		21	A14 -> A13

Step 6: Establish all α – order fuzzy logical relationships groups

Based on [2] all the fuzzy relationships having the same fuzzy set on the left-hand side or the same current state can be put together into one fuzzy relationship group. The fuzzy logical relationship as the same are counted only once. Thus, from Table 3 and based on Definition 4, we can obtain seven first – order fuzzy logical relationship groups shown in Table 4.

Table 4. The complete results for the first - order fuzzy relationship groups.

No group	Fuzzy relation groups	No group	Fuzzy relation groups
1	A1 -> A1, A2	8	A9 -> A11
2	A2 -> A3	9	A11 -> A8
3	A3 -> A5, A4	10	A8 -> A5
4	A5 -> A4, A5, A3	11	A10 -> A12
5	A4 -> A6, A7	12	A12 -> A13
6	A6 -> A7	13	A13 -> A14
7	A7 -> A9, A10	14	A14 -> A14, A13

Step 7: Calculate and defuzzify the forecasted values

Calculate the forecasted output at time t by using the following principles:

Principle 1: If the fuzzified value of year t-1 is A_j and there is only one fuzzy logical relationship in the fuzzy relationship group whose current state is A_j, shown as follows: $\mathrm{A_j} \rightarrow \mathrm{A_k}$; then the forecasted value of year t forecasted = m_k; where m_k is the midpoint of the interval u_k and the maximum membership value of the fuzzy set A_k occurs at the interval u_k.

Principle 2: If the fuzzified value of year t -1 is A_j and there are the following fuzzy logical relationship group whose current state is A_j, shown as follows:

$$\mathrm{A_j} \rightarrow \mathrm{A_{i1}}, \mathrm{A_{i2}}, \mathrm{A_{ip}}$$

then the forecasted value of year t is calculated as follows: $forecasted = \frac{\sum_{k=1}^{p} m_{ik}}{p}$; where m_{i1}, m_{i2}, m_{ik} are the middle values of the intervals u_{i1}, u_{i2} and u_{ik} respectively, and the maximum membership values of A_{i1}, A_{i2},..., A_{ik} occur at intervals u_{i1}, u_{i2}, u_{ik}, respectively.

Based on two these rules, we complete forecasted results for all years of enrolment which are shown in Table 5.

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

Year	Actual	Fuzzified	Results
1971	13055	A1	Not forecasted
1972	13563	A1	13660
1973	13867	A2	13660
1974	14696	A3	14795
------	-----	----	-----
1989	18970	A13	18832
1990	19328	A14	19332
1991	19337	A14	19082
1992	18876	A13	19082

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 (11) as follows:

$$\mathrm{R}MSE = \sqrt{\frac{1}{n}\sum_{t=\alpha}^{n} (F_t - R_t)^2} \quad (11)$$

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.

3.2. FTS Forecasting Model Combining the K-means and HSA

The goal of this subsection is that we present the hybrid FTS forecasting model by combining K-means algorithm for partition data set into the unequal lengths of intervals with HSA in Subsection 2.3. The main purpose of this algorithm is to adjust the initial intervals length with an intent to obtain the optimal intervals without increasing the number of intervals. The detailed descriptions of the hybrid forecasting model are given as follows. In proposed model, each Harmony 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 on historical data Y (t) be x_0 and x_n, respectively. Each harmony i is a vector consisting of n-1 elements x_k where 1≤ k ≤ n-1 and $x_k \leq x_{k+1}$. Based on these n-1 elements, define the n intervals as $u_1 = [x_0, x_1]$, $u_2 = [x_1, x_2]$,,..., $u_i = [x_{k-1}, x_k]$,... and $u_n = [x_{n-1}, x_n]$, 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 x_k $(1 \leq k \leq n-1)$ arranges in an ascending order.

The parameters of the harmony search algorithm that are supposed to be defined in this section are harmony memory size (HMS), i.e., the number of solution vectors or rows in the harmony memory matrix; harmony memory considering rate (HMCR); pitch adjusting rate (PAR); bandwidth (BW); and the number of iterations [23]. The algorithm generates random solution vectors (harmonies) HMS times and puts them in the HM matrix, specified by formula (2). Improvise a new harmony: Once the harmony memory matrix is initialized, the algorithm starts the first iteration by improvising a new harmony. X = (x_1, x_2,..., x_{n-1}) is a new solution vector that is constructed based on three rules:

1) Memory consideration with probability HMCR.

2) Pitch adjusting with probability PAR.

3) Random selection with probability (1-HMCR).

The complete steps of the hybrid forecasting model are presented in Algorithm 1.

Algorithm 1: FTS-K_HSA algorithm

Input: *time series data*

Output: *the RMSE, forecasted value and optimum intervals*

1) Select the initial set of intervals by applying K - mean algorithm and use forecasting steps in Subsection 3.2 to get the initial forecasting accuracy (MSE).

2) Initialize all parameters as follows:

HMCR = 0.99, PAR = 0.5, BW = 1.

Lb = x_0= 13000; Ub = x_n = 20000;

Maximum number of improvisations is 100;

Number of Harmonys HMS = 30,

Harmonies are Initialized $x_0 + Rand() * (x_n - x_0)$;

3) While the stop condition (maximum number of improvisations or minimum RMSE criteria) is not satisfied do

For Harmony i, (1≤ $i \leq$ HMS) do

i. Define fuzzy sets according to all intervals defined by Harmony i.

ii. Fuzzify all historical data by Step 4 in Subsection 3.1.

iii. Create all $\alpha -$ order fuzzy relationships by Step 5 in Subsection 3.1.

iv. Establish all $\alpha -$ order fuzzy relationships groups by Step 6 in Subsection 3.1.

v. Calculate forecasting values by Step 7 in Subsection 3.1.

vi. Compute the F (X) = RMSE values for Harmony i based on formula (11).

End for

For each Harmony i ∈ HMS

vii. Update HM (3) i according to the RMSE values mentioned above.

viii. Update the RMSE = F $(X)_{HMS\times 1}$values.

End for

End while

4. Experimental Results

In this study, the performance of the proposed model is evaluated based on two different datasets consisting of enrollments data of University of Alabama [2] the period from 1971 to 1992, and vehicle road accidents dataset [1] period between 1974 and 2004.

4.1. Forecasting the Road Traffic Accidents

In this section, the proposed model is also applied to forecast the road traffic accidents in Belgium [25] from 1974 to 2004 and there is made a comparison of the forecasting results with the previous works [25-27]. A comparison of the forecasting results using RMSE (10) is shown in Table 6. It is evident that the proposed method gets better forecasting results than the forecasting models above. More detailed comparison, for the same number of 13, the proposed model obtains the smallest RMSE value of 4.9 among two models [27] using the 3rd - order FTS. Besides, the proposed model also has far smaller RMSE value of 17.8 than the models [25, 26] based on first - order FTS with different number of intervals.

Table 6. A comparison of the forecasting results between proposed model and various models based on first – order and three - order FTS.

Year	Actual data	Model [25]	Model [26]	Model [27]	Proposed model	
1974	1574	---	----	----	---	----
1975	1460	1506	1458	----	1551.8	----
1976	1536	1453	1467	----	1444.3	----
1977	1597	1598	1606	1594	1596	1598.5
1978	1644	1584	1592	1643	1628.8	1643.66
----	-----	-----	-----	-----		
2003	1035	970	1097	1036	1008.09	1025.5
2004	953	970	929	954	986.25	958
RMSE		41.61	37.66	19.2	17.8	4.9

4.2. Forecasting Enrolments of University of Alabama

In this section, the proposed forecasting model is applied for forecasting enrolments from yearly observations [6]. To show the performance of the proposed forecasting model based on the first - order FTS under the number of intervals of 14, four forecasting models presented in articles [1, 2, 5, 8] are selected for the purpose of comparison. The RMSE value is obtained from the proposed forecasting model, as listed in Table 7 is far smaller than that of all the existing forecasting models mentioned above. Furthermore, to verify the forecasting performance of the proposed model under different number of intervals and different high - order FRGs, five FTS models presented in articles [3, 4, 6, 16, 28] which are examined and compared. The forecasted accuracy of the proposed method is estimated by using the RMSE (11). A comparison of the forecasting accuracy with various orders and number of intervals of 14 among the models [3, 4, 6, 16, 28] and the proposed model are shown in Table 8. From Table 8, it can be seen that proposed model has a RMSE value is 11.73 which is the lowest among all forecasting models compared.

In addition, the proposed model is also compared with it counterparts [3, 4, 16, 28] based on various high – order fuzzy relationships under seven. The details of comparison are shown in Table 9.

Table 7. A comparison of the forecasted results of proposed model with its counterparts based on first-order FTS under number of intervals of 14.

Year	Actual data	Model [8]	Model [2]	Model [5]	Model [1]	Proposed model	
						FTS-KM	*FTS_K-HSA*
1971	13055	-	-	-	-	-	-
1972	13563	14000	14000	14000	13714	13660	13558.66
1973	13867	14000	14000	14000	13714	13660	13751.5
1974	14696	14000	14000	14000	14880	14795	14667.5
---	----	----	----	----	----	----	------
1988	18150	16813	16833	17500	18055	18035	18065
1989	18970	19000	19000	19000	18998	18832	18893
1990	19328	19000	19000	19000	19300	19332	19289.75
1991	19337	19000	19000	19500	19149	19082	19289.75
1992	18876	19000	19000	19149	19014	19082	19053.88
RMSE		650.4	638.4	476	187.9	161.6	85.1

Table 8. A comparison of the forecasted results of the proposed model with its counterparts based on high – order FTS under different number of intervals.

Years	Actual data	Model [6]	Model [3]	Model [4]	Model [16]	Model [28]	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	N/A	N/A	N/A	N/A	N/A	N/A
1974	14696	N/A	N/A	N/A	N/A	N/A	N/A
1975	15460	15500	N/A	N/A	N/A	N/A	N/A
1976	15311	15468	15500	N/A	N/A	N/A	15303.83
1977	15603	15512	15500	N/A	N/A	N/A	15601.83
1978	15861	15582	15500	N/A	N/A	N/A	15851
1979	16807	16500	16500	16846	N/A	N/A	16802.5
1980	16919	16361	16500	16846	16890	16920	16920.5

Years	Actual data	Model [6]	Model [3]	Model [4]	Model [16]	Model [28]	Proposed model
1981	16388	16362	16500	16420	16395	16388	16391
-------	------	-------	------	------	-----	------	------
1990	19328	19382	19500	19334	19337	19338	19331.92
1991	19337	19487	19500	19334	19337	19335	19331.92
1992	18876	18744	18500	18910	18882	18882	18846.5
RMSE		365.65	294.44	33.18	15.3	13.15	11.73

Table 9. A comparison of the RMSE value between proposed model its counterparts under different number of orders with the number of interval is 7.

Models	Number of order of forecasting models							
	2	3	4	5	6	7	8	9
Model [3]	298.48	294.44	298.96	307.47	313.39	322.58	319.65	320.61
Model [4]	260.45	176.42	178.91	158.06	164.26	164.22	149.62	136.87
Model [16]	259.08	177.89	152.55	153.41	153.85	143.7	130.79	134.06
Model [28]	139.98	176.6	141.97	142.71	149.25	135.95	121.56	123.49
Proposed model	129.7	55	38.84	37.44	44.12	34.58	32.3	32.8

From Table 9, it can be seen that the accuracy of the proposed model is improved significantly. Particularly, the proposed model gets the lower RMSE value than four models presented in articles [3, 4, 16, 28]. To sum up, These finding suggests that the proposed model is able to provide effective forecasting capability based on first – order and high – order FTS model with the different number of orders and the different number of intervals.

5. Conclusion

In this paper, a hybrid FTS forecasting model based on combining K-means and Harmony search algorithm for forecasting the various problems. The advantages of the proposed model are that it combines K-means and Harmony search algorithm to get the optimum partition of the intervals for increasing the forecasting accuracy rates. These combination of optimum techniques always leads to the development of new architecture, which should be more advantageous, providing robust, cost effective, and approximate solution in comparison to conventional techniques. From the comparison results in Tables 6-9, it can be seen that the proposed model outperforms previous forecasting models in the training phase based on both first – order and high – order FTS.

References

[1] Chen S-M, & Chung, N.-Y (2006a). Forecasting enrollments using high-order fuzzy time series and genetic algorithms. Research Articles International Journal of Intelligent Systems, 21, 485–501.

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

[3] Chen S M (2002). Forecasting enrolments based on high-order fuzzy time series. Cybernetics and Systems, 33 (1), 1-16.

[4] Chen S-M & Chung N-Y (2006b). Forecasting enrollments of students by using fuzzy time series and genetic algorithms. International Journal of Intelligent Systems, 17, 1-17.

[5] Huarng K (2001). Effective lengths of intervals to improve forecasting in fuzzy time series. Fuzzy Sets and Systems, 123 (3), 387-394.

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

[7] Song Q & Chissom. B S (1993a). Fuzzy time series and its models. Fuzzy Sets and Systems, 54 (3), 269-277.

[8] Song Q & Chissom. B S (1993b). Forecasting enrolments with fuzzy time series - Part I. Fuzzy Sets and Systems, 54 (1), 1-9.

[9] Song Q & S. Chissom. B (1994). Forecasting enrolments with fuzzy time series - part II. Fuzzy Sets and Systems, 62, 1-8.

[10] Yu H K (2005). A refined fuzzy time-series model for forecasting. Physical A: Statistical Mechanics and Its Applications, 346 (3-4), 657-681.

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

[12] Chen T-L, Cheng C-H & Teoh H-J (2008). High-order fuzzy time-series based on multi-period adaptation model for forecasting stock markets. Physical A: Statistical Mechanics and Its Applications, 387, 876–888.

[13] Chu H-H, Chen T-L, Cheng C-H & Huang C-C (2009). Fuzzy dual-factor time- series for stock index forecasting. Expert Systems with Applications, 36, 165–171.

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

[15] al L-Y H e (2010). Temperature prediction and TAIFEX forecasting based on fuzzy relationships and MTPSO techniques. Expert Syst Appl, 37, 2756–2770.

[16] I. H. Kuo e a (2009). An improved method for forecasting enrolments based on fuzzy time series and particle swarm optimization. Expert Systems with Applications, 36, 6108–6117.

[17] Kuo. I-H, Horng. S-J, Che. Y-H, R-S R, Kao. T-W, R.-J. C. J-L L, et al. (2010). Forecasting TAIFEX based on fuzzy time series and particle swarm optimization. Expert Systems with Applications, 2 (37), 1494–1502.

[18] Pritpal. S & Bhogeswar B (2013). High-order fuzzy-neuro expert system for time series forecast-ing. Knowl-Based Syst, 46, 12-21.

[19] L. Al - Mataneh e a (2014). Development of temperature based weather forecasting model using neural network and fuzzy logic. International Journal of Multimedia and Ubiquitous Engineering, 9 (12), 343-366.

[20] S. Askari & Montazerin N (2015). A high-order multi-variable fuzzy time series forecasting algorithm based on fuzzy clustering. Expert Systems with Applications, 42, 2121 -2135.

[21] B. Sun, H. Guo, H. R. Karimi & Y. Ge S X (2015). Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series. Neuro computing, 151, 1528-1536.

[22] Gersho A & Gray R M (1992). Vector Quantization and Signal Compression. KAP.

[23] Geem Z W, Kim J H & Loganathan G V (2001). A new heuristic optimization algorithm: Harmony search. Simulation, 76, 60–66.

[24] Fesanghary M, Mahdavi M, Minary M & Alizadeh (2008). Y. Hybridizing harmony search algorithm with sequential quadratic programming for engineering optimization problems. Comput Methods Appl Mech Energy, 197, 3080–3091.

[25] Uslu V R, et al (2013). A fuzzy time series approach based on weights determined by the number of recurrences of fuzzy relationships. Swarm Evol Comput, doi: 10.1016/j. swevo.2013.10.004.

[26] Bas E, et al (2014). A modified genetic algorithm for forecasting fuzzy time series. Appl Intell, 453-463 DOI 10.1007/s10489-014-0529-x.

[27] Yusuf S M, et al. (2015). A Novel Hybrid fuzzy time series Approach with Applications to Enrollments and Car Road Accident. International Journal of Computer Application, 192 (2), 37-44.

[28] Y.-L. Huang e a (2011). a hybrid forecasting model for enrolments based on aggregated fuzzy time series and particle swarm optimization. Expert Systems with Applications, 7 (38), 8014–8023.

[29] MacQueen J B. "Some methods for classication and analysis of multivariate observations" in: Proceedings of the Fifth Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, CA1967. p. 281-297.

[30] Abhishekh, S. S. Gautam and S. R. Singh, A refined method of forecasting based on high-order intuitionistic fuzzy time series data, Progr. Artif. Intell. 7 (4) (2018) 339–350.

[31] Abhishekh and S. Kumar, A computational method for rice production forecasting based on high-order fuzzy time series, Int. J. Fuzzy Math. Archive 13 (2) (2017) 145–157.

[32] S. Xian, J. Zhang, Y. Xiao and J. Pang, A novel fuzzy time series forecasting method based on the improved artificial fish swarm optimization algorithm, Soft Comput. 22 (12) (2018) 3907–3917.

[33] R. Efendi, N. Arbaiy and M. M. Deris, A new procedure in stock market forecasting based on fuzzy random auto-regression time series model, Inform. Sci. 441 (2018) 113–132.