



Pythagorean fuzzy time series model based on Pythagorean fuzzy c-means and improved Markov weighted in the prediction of the new COVID-19 cases

Sidong Xian¹ · Yue Cheng¹

Accepted: 8 September 2021 / Published online: 6 October 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Time series is an extremely important branch of prediction, and the research on it plays an important guiding role in production and life. To get more realistic prediction results, scholars have explored the combination of fuzzy theory and time series. Although some results have been achieved so far, there are still gaps in the combination of n -Pythagorean fuzzy sets and time series. In this paper, a pioneering n -Pythagorean fuzzy time series model (n -PFTS) and its forecasting method (n -IMWPFM) are proposed to employ a n -Pythagorean fuzzy c-means clustering method (n -PFCM) to overcome the subjectivity of directly assigning membership and non-membership values, thus improving the accuracy of the partition the universe of discourse. A novel improved Markov prediction method is exploited to enhance the prediction accuracy of the model. The proposed prediction method is applied to the yearly University of Alabama enrollments data and the new COVID-19 cases data. The results show that compared with the traditional fuzzy time series forecasting method, the proposed method has better forecasting accuracy. Meanwhile, it has the characteristics of low computational complexity and high interpretability and demonstrates the superiority of this model from a realistic perspective.

Keywords n -Pythagorean fuzzy sets · n -Pythagorean fuzzy time series · n -Pythagorean fuzzy c-means clustering · COVID-19 · Improved Markov weighted

1 Introduction

Using the known to derive the unknown, that is prediction, has important guiding significance for production and life, specifically collecting long-term weather data to predict whether a ship is seaworthy, looking at seasonal variations in climate to predict suitability for farming, using stock trading data to predict the timing of purchases. Accordingly, the number of COVID-19 cases in a period of time is used to predict the number of new cases in a period of time in the future. In short, predictions are everywhere and extremely useful. Therefore, finding the law of things running from the complicated reality has become a game that people are diligently seeking. Fuzzy time series is an important branch of

prediction methods, and with the passage of time, researchers contributed more and more literature to expand it Song and Chissom (1993), Egrioglu et al. (2013), Dincer and Akkuş (2018), Singh (2017), and Xian et al. (2018c, 2020).

After Zadeh (1965) completed the theoretical work of fuzzy sets, which provided an unprecedented idea for dealing with uncertain and fuzzy linguistic variables, Song and Chissom (1993, 1965a, b) developed the first fuzzy time series model and prediction method based on this. Taking it as a starting point, the research on time series has been fruitful in recent years and in various forms. Egrioglu et al. (2013) proposed a hybrid fuzzy time series method, which uses the fuzzy c-means clustering method to achieve the fuzzy data. Vovan and Ledai (2019) employs an automated algorithm to determine the appropriate number of clusters. Duru and Bulut (2014) proposed a novel clustering method named histogram damping partition (HDP), which can be used for many different kinds of fuzzy time series models at the clustering stage. Dincer and Akkuş (2018) proposed a fuzzy time series method based on robust clustering to optimize the fuzzification step. And by the same token, Dincer (2018) also used

✉ Sidong Xian
sidx@163.com

¹ Key Laboratory of Intelligent Analysis and Decision on Complex Systems, Chongqing University of Posts and Telecommunications, Chongqing 400065, People's Republic of China

fuzzy c -regression to achieve this goal. Bionic algorithms are also popular. Bas et al. (2014) employed a modified genetic algorithm for forecasting fuzzy time series. Deng et al. (2015) used the heuristic Gaussian cloud transformation algorithm to extract the uncertain numerical time series into Gaussian clouds to construct the dataset. Singh (2017) introduced a new fuzzy time series (FTS) forecasting model, which tries to deal with four major associated issues. Cai et al. (2015) presented a fuzzy time series model combined with ant colony optimization (ACO) and auto-regression. Xian et al. (2020) proposed a new fuzzy time series forecasting model which considers a hybrid wolf pack algorithm (HWPA) and an ordered weighted averaging (OWA) aggregation operator for a fuzzy time series.

As an alternative to fuzzy logic relations, neural networks are now attracting the attention of researchers. Based on the single multiplication neuron model, Aladag (2013) proposed a new artificial neural network (ANN) fuzzy time series method. Abd Rahman et al. (2015) employed an artificial neural network method to construct logical relations thus presented a time series model in predicting the air pollution index (API) from three different stations. Egrioglu et al. (2019) proposed an intuitionistic high-order fuzzy time series forecasting method based pi-sigma artificial neural networks Shin and Ghosh (1991) trained by artificial bee colonies. Passionate about the ANN algorithm, Egrioglu et al. (2020) also proposed the Picture fuzzy time series model and also used ANN to replace the construction of fuzzy logic relations. However, because of the inexplicability of artificial neural networks, many scholars still prefer traditional prediction methods, such as Tsaur (2012) and Efendi et al. (2013) using the Markov weighted method for prediction, and Li and Cheng (2010), Li et al. (2010) is based on hidden Markov chains and long-term models.

In addition to the partition of the universe and the construction of fuzzy relations, the definition of fuzzy intervals is also the focus of research on fuzzy time series. The fuzzy sets employed by Song and Chissom (1993) are not complete enough to express uncertain information. Therefore, many scholars are devoted to the expansion and improvement in fuzzy set theory. When researchers express membership degrees in a fuzzy set, they do not always express non-membership degrees as a complement of 1. Aimed at the situation, Atanassov (1986) proposed intuitionistic fuzzy sets. Intuitionistic fuzzy sets give μ (membership degrees), ν (non-membership degrees), and $\mu + \nu \leq 1$. On the basis of this work, Castillo et al. (2007) first integrated intuitionistic fuzzy set inference into the analysis of time series and initially established the prediction model of intuitionistic fuzzy time series. Olej and Hájek (2010) have designed an intuitionistic fuzzy inference system for the ozone time series prediction. Kumar and Gangwar (2015) and Fan et al. (2017) used intuitionistic fuzzy time series to deal with the uncertainty in

time series prediction. Abhishekh Gautam and Singh (2018) proposed a high-order intuitionistic fuzzy time series model based on score function and fine prediction methods. Bas et al. (2020) proposed a new intuitionistic fuzzy regression functions approach for forecasting purposes.

However, intuitionistic fuzzy sets also have limitations. Take the COVID-19 patient as an example, if the probability of a suspected patient being diagnosed using method A is 0.8, the probability of using method B to be healthy is 0.6. In response to this situation, the intuitionistic fuzzy time series is helpless. However, the Pythagorean fuzzy sets proposed by Yager (2013) can effectively deal with this kind of situation. Yager (2014) and Yager and Abbasov (2013) gave the details of the theory and Xian et al. (2018a, b) gave some applications for it. On the basis of previous work, Bryniarska (2020) presents deductive theories of n -Pythagorean fuzzy sets, which makes the restriction of membership degree and non-membership degree smaller and can reflect uncertainty more truly. For a situation similar to the above example and the intuitionistic fuzzy time series cannot cope with, this paper proposes an improved Markov weighted n -Pythagorean fuzzy time series forecasting method based on n -Pythagorean fuzzy c -means (n -IMWPFM). The prediction steps of this method are divided into two steps. First, a n -Pythagorean fuzzy c -means clustering method is proposed to fuzzy the real data points, and the training sets constructed by membership and non-membership degrees are obtained. Then, the improved Markov weighting method is employed to complete the defuzzification work. The detailed contributions of the paper are:

- A novel n -Pythagorean fuzzy time series is proposed, and a forecasting method based on it is given. The proposed model can better express the uncertainty of time series data, and its prediction method has better prediction accuracy.
- A novel n -Pythagorean fuzzy c -means clustering (n -PFCM) method is proposed, which can be exploited to obtain the n -Pythagorean fuzzy number and achieve better universe segmentation.
- An improved Markov Weighted prediction algorithm is proposed and applied to the defuzzification of the n -Pythagorean fuzzy time series. This improves the accuracy of the prediction result while avoiding a large number of identical results.
- The models and methods presented in this paper are employed to predict the new COVID-19 cases data. The prediction method can provide valuable reference opinions for the control of the COVID-19 epidemic in the world.

The rest of this paper is organized as follows: Sect. 2 briefly reviews several basic concepts concerning n -

Pythagorean fuzzy sets, fuzzy c-means clustering, and Markov weighted matrix. Section 3 proposes two novel methods from the partition of the universe of discourse and defuzzification, one is the n -Pythagorean fuzzy c-means clustering method and the other is an improved Markov weighted prediction method. In Sect. 4, this paper gives the relevant definition of the n -Pythagorean fuzzy time series and gives the steps of the forecasting method in detail. Section 5 discusses two numerical examples based on the above prediction method, one is yearly University of Alabama enrollments data and the other is the new COVID-19 cases data in the U.S. from January 23, 2020, to April 19, 2021 (Dong et al. 2020); the accuracy of the proposed method is finally verified through MSE, RSME, and SMAPE. Conclusions and future research directions are given in Sect. 6.

2 Preliminaries

2.1 n -Pythagorean fuzzy sets (n -PFSs)

Yager (2013) pioneered the concept of the Pythagorean fuzzy set and provided comprehensive operation and application examples of the sets (Yager 2014). In order to further expand the applicability of Pythagorean fuzzy sets in practical applications, Anna proposed n -Pythagorean fuzzy sets.

Definition 1 (Bryniarska 2020, n -Pythagorean fuzzy sets) Let X be an universe of discourse, a n -Pythagorean fuzzy set P can be defined as shown below:

$$P = \{ \{x, \mu_p(x), \nu_p(x)\} | x \in X \}, \tag{1}$$

Among them:

$$\begin{aligned} \mu_p(x) : X &\rightarrow [0, 1], x \in X \rightarrow \mu_p(x) \in [0, 1]; \\ \nu_p(x) : X &\rightarrow [0, 1], x \in X \rightarrow \nu_p(x) \in [0, 1]; \end{aligned}$$

and

$$0 \leq \mu_p(x)^n + \nu_p(x)^n \leq 1, x \in X;$$

in addition

$$\pi_p(x)^n = 1 - \mu_p(x)^n - \nu_p(x)^n, x \in X;$$

where $\mu_p(x), \nu_p(x), \pi_p(x)$ represents the membership, non-membership and hesitation of the element x belonging to X , respectively. Remember $p = (\mu_p(x), \nu_p(x))$ as a n -Pythagorean fuzzy number.

2.2 Fuzzy c-means clustering (FCM)

Fuzzy c-means is an unsupervised machine learning method of clustering which allows one piece of data to belong to two or more clusters. The fuzzy theory is introduced into the c-means clustering algorithm proposed by Bezdek et al. (1984) to obtain the FCM, which realizes segmentation by calculating the quantitative relationship between data and the membership degree of clustering centers and relies on iteration to minimize the objective function J_m .

Definition 2 (Dunn 1973) The objective function J_m is denoted and defined by the following function:

$$J_m = \sum_i^N \sum_j^c \mu_{ij}^m d(X_i, C_j). \tag{2}$$

where d is Euclidean distance measure between different cluster center (C_j) and data points (X_i) and μ_{ij} is the membership value of i th data (X_i) in j th cluster. c is the number of clusters; N is the number of data points. m is a constant representing which means a fuzzifier ($m = 2$).

Minimization of J_m is based on suitable selection of μ_{ij} (membership matrix) and C_j using an iterative process through the following equation:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c (\| \frac{X_i - C_j}{X_i - C_k} \|)^{\frac{2}{m-1}}}. \tag{3}$$

$$C_j = \frac{\sum_{i=1}^N \mu_{ij}^m X_i}{\sum_{i=1}^N \mu_{ij}^m}. \tag{4}$$

2.3 Markov chain

The Markov chain is a random process that depends on conditional probabilities, and it converges to a set of vectors under certain conditions (Ross 2013).

Definition 3 (Ross 2013, Markov chain) A Markov chain is a set of discrete random variables with Markov properties. Specifically, for a set of random variables $X = X_n : n > 0$ with a one-dimensional countable set as the exponent set in the probability space $(\Omega, \Gamma, \mathbb{P})$, if the values of the random variables are all in the countable set: $(X = s_i, s_i \in s)$ and the conditional probability of the random variable satisfies the following relationship:

$$p(X_{t+1}|X_t, \dots, X_1) = p(X_{t+1}|X_t), \tag{5}$$

then X is called a Markov chain, the countable set $s \in \mathbb{Z}$ is called the state space, and the value of the Markov chain in the state space is called the state.

Definition 4 (Ross 2013, *Transition Probability*) Let $\{X_n, n \geq 0\}$ be a given Markov chain, for any $m \geq 0, n \geq 1, i, j \in E$, if

$$p_{ij}(m, m+n) = p\{X_{m+n} = j | X_m = i\}, \quad (6)$$

then call the probability that the state of the chain is i when the chain is m , and then transfer to j through n steps, referred to as the n -step transition probability; especially, the one-step transition probability is $p_{ij}(m, m+1)$.

Definition 5 (Ross 2013, *Markov Transition Probability Matrix*) Let A be a $n \times n$ matrix,

$$A = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix},$$

$$p_{ij} \in A(i, j = 1, 2, 3, \dots, n), s.t.$$

$$p_{ij} \in [0, 1], \quad \sum_{j=1}^n p_{ij} = 1,$$

where p_{ij} is expressed by probability, which can be transferred to each other under certain conditions; then, A is called Markov transition probability matrix.

Tsaur (2012) and Efendi et al. (2013) have completed the pioneering work of introducing Markov weighting matrices into time series. The specific steps are introduced in detail in Sect. 3.

3 n -Pythagorean fuzzy c-means clustering (n-PFCM) and improved Markov weighted prediction method

3.1 n -Pythagorean fuzzy c-means clustering

In practice, data points usually appear as real numbers, while fuzzy sets usually require researchers to develop. A fuzzy clustering algorithm, as a fast and effective method to convert data points from real numbers to fuzzy numbers, has attracted much attention. In the following decades, with the development of fuzzy set theory, the fuzzy c-means method has been continuously improved by scholars in different fields for different purposes. An intuitionistic fuzzy c-means algorithm to the cluster of IFSs is developed by Xu and Wu (2010).

Zhao et al. (2014) proposed an improved FCM segmentation method based on the intuitionistic fuzzy set (IFS), to resolve the problem that the FCM algorithm cannot retain image details well and it is difficult to segment small regions. Verma et al. (2016) presented an improved intuitionistic fuzzy c-means (IIFCM), which considers the limited spatial information in an intuitionistic fuzzy way. Thong and Son (2016) proposed a picture fuzzy clustering method inspired by the fuzzy c-means algorithm.

In view of this, this paper proposes a simpler and faster method to obtain a n -Pythagorean fuzzy set, which is called n -Pythagorean fuzzy c-means clustering. In the proposed n -Pythagorean fuzzy c-means clustering algorithm, the objective function that is supposed to be minimized contains two terms: one is modified objective function of the conventional FCM using n -PFS and the other is to exploit the method proposed by Thong and Son (2016) to optimize the objective function.

3.1.1 A novel fuzzy c-means clustering based on n -Pythagorean fuzzy set

In order to incorporate the fuzzy attributes of n -Pythagorean into the conventional fuzzy clustering algorithm, this paper defines the clustering center, membership degree, hesitation degree and n -Pythagorean fuzzy c-means clustering model as follows.

Definition 6 (*The cluster center of n-PFCM*) Let X be an universe of discourse, then the cluster center C_j^* can be defined as shown below:

$$C_j^* = \frac{\sum_{i=1}^N \mu_{ij}^{*m} X_i}{\sum_{i=1}^N \mu_{ij}^{*m}}; \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, c. \quad (7)$$

Among them, μ_{ij}^* is the degree of membership after integrating the degree of n -Pythagorean membership and hesitation, which is calculated as follows:

$$\mu_{ij}^* \rightarrow \left(\mu_{ij}^n + \pi_{ij}^n \right)^{\frac{1}{n}}. \quad (8)$$

where m is a constant representing which means a fuzzifier, n is a variable exponent and $n \in \mathbb{Z}$, X_i are the real data points, μ_{ij} and π_{ij} are the degree of n -Pythagorean membership and hesitation.

Definition 7 (*The n-Pythagorean membership of X_i*) Let X be a universe of discourse, the n -Pythagorean membership

(μ_{ij}) of X_i can be defined as shown below:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\left\| \frac{X_i - C_j^*}{X_i - C_k^*} \right\| \right)^{\frac{2}{m-1}}}. \tag{9}$$

where u_{ij} is membership of X_i in j th cluster center, π_{ij} is hesitation of X_i in j th cluster center, C_j^* is cluster center which obtained by n PFCM.

Definition 8 (The n -Pythagorean hesitation of X_i) Let X be a universe of discourse, the n -Pythagorean hesitation (π_{ij}) of X_i can be defined as shown below:

$$\pi_{ij}^n = \left((\mu_{ij}^n + v_{ij}^n)^\alpha - (\mu_{ij}^n)^\alpha \right)^{\frac{1}{\alpha}}. \tag{10}$$

Among them, the value of v_{ij} is calculated as follows:

$$v_{ij}^n = 1 - \mu_{ij}^n - \pi_{ij}^n = 1 - \mu_{ij}^n - \left((\mu_{ij}^n + v_{ij}^n)^\alpha - (\mu_{ij}^n)^\alpha \right)^{\frac{1}{\alpha}}. \tag{11}$$

where n is a variable exponent and $n \in \mathbb{Z}$, $\alpha \in (0, 1)$, which is generally selected as 0.8, u_{ij} is n -Pythagorean membership of X_i in j^{th} cluster center, π_{ij} is n -Pythagorean hesitation of X_i in j^{th} cluster center, and C_j^* is cluster center which obtained by n -PFCM.

Remark 1 If the value of n in n -Pythagorean membership degree changes, then n -Pythagorean membership degree also changes as follows:

- (1) If $n = 1$, n -Pythagorean membership reverts to intuitionistic membership.
- (2) If $n = 2$, n -Pythagorean membership reverts to Pythagorean membership.
- (3) If $n = 3$, n -Pythagorean membership reverts to Fermatean membership (Senapati and Yager 2020).

Model 1 (The n -Pythagorean fuzzy c -means clustering model)

Let $P = \{ \{ X_i, \mu_{ij}(X_i), v_{ij}(X_i) \} \}$, represents the n -PFSs of the original data to be divided into c clusters, where X_i represents the true value at the i th data. Then, the objective function J_{n-PFCM} of the n -PFSs model is given as:

$$J_{n-PFCM} = \sum_i^n \sum_j^c \mu_{ij}^* d(X_i, C_j^*), 1 < m < \infty. \tag{12}$$

where C_j^* is the j th cluster center, μ_{ij} is the degree of n -Pythagorean memberships, $d(\cdot)$ is the Euclidean distance, m is a fuzzifier.

$$s.t. \begin{cases} \mu_{ij}, v_{ij}, \pi_{ij} \in [0, 1], & i = 1, 2, \dots, n; j = 1, 2, \dots, c. \\ \mu_{ij}^n + v_{ij}^n + \pi_{ij}^n = 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, c. \\ \sum_{j=1}^c \mu_{ij}^* = 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, c. \\ 0 < \sum_{j=1}^c \mu_{ij}^{*m} < 1, & i = 1, 2, \dots, n; j = 1, 2, \dots, c. \end{cases} \tag{13}$$

Employing Eqs. (7)–(11) update the cluster center and update the member matrix at the same time. In each iteration, the cluster center and the member matrix are updated, and the following conditions need to be met eventually.

$$\max \{ | \mu_{ij}^{*t+1} - \mu_{ij}^{*t} | + | v_{ij}^{*t+1} - v_{ij}^{*t} | \} < \varepsilon. \tag{14}$$

In this paper, the value of ε takes 0.05, the fuzzifier m takes 2. The pseudocode of the n -Pythagorean fuzzy c -means clustering algorithm is summarized in Algorithm 1.

Algorithm 1 Pseudocode for the n -Pythagorean fuzzy c -means clustering.

Input:

Initialize the $n, c, m = 2, \alpha = 0.8$ and $\varepsilon = 0.05$.

Initialize the C_j^*, μ_{ij}, v_{ij} according to Eq. (13).

Output:

$C_j^*, \mu_{ij}, v_{ij}, \pi_{ij}$.

1. **While** the termination condition is not satisfied,
 2. Calculate the cluster centers c_j as Eq. (7).
 3. Update the μ_{ij}, v_{ij} , and π_{ij} as Eqs. (9–11).
 4. **for** assign the μ_{ij} by Eq. (8).
 5. Update the π_{ij} by Eq. (10).
 6. Update the v_{ij} by Eq. (11).
 7. **end for**
 8. **end while**
 9. **Return** the C_j^*, μ_{ij}, v_{ij} and π_{ij} .
-

3.1.2 Optimized n -Pythagorean fuzzy c -means clustering model

Inspired by Thong and Son (2016), model 1 is optimized and model 2 is proposed in this paper, which transforms the objective function into a convex function and obtains the optimal solutions for both affiliation and non-affiliation by Lagrangian methods.

Model 2 (The optimized n -Pythagorean fuzzy c -means clustering model) Supposing that X be a real number set. Let

$P = \{ \{ X_i, \mu_{ij}(X_i), v_{ij}(X_i) \} \}$, represents the n PFS of the X to be divided into c clusters, where X_i represents the true value at the i th data. Then, the objective function J_{nOPFCM}

of the n PFSs model is given as:

$$J_{nOPFCM} = \sum_i^n \sum_j^c (2\mu_{ij}^n)^m d(X_i, C_j^*) + \sum_{i=1}^n \sum_{j=1}^c n v_{ij}^n \ln v_{ij}, 1 < m < \infty. \tag{15}$$

where C_j^* is the j th cluster center, μ_{ij} is the degree of n -Pythagorean memberships, v_{ij} is the degree of n -Pythagorean non-memberships $d(\cdot)$ is the Euclidean distance, m is a fuzzifier.

$$s.t. \begin{cases} \mu_{ij}, v_{ij}, \pi_{ij} \in [0, 1], \\ \mu_{ij}^n + v_{ij}^n + \pi_{ij}^n = 1, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, c. \\ \sum_{j=1}^c 2\mu_{ij}^n = 1, \end{cases} \tag{16}$$

Theorem 1 Let the objective function J_{nOPFCM} be minimized, s.t. Eq. (16). Then, the optimal solutions of the Model 2 are:

$$C_j^* = \frac{\sum_{i=1}^N (2\mu_{ij}^n)^m X_i}{\sum_{i=1}^N (2\mu_{ij}^n)^m}. \tag{17}$$

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c 2^{\frac{1}{n}} \left(\frac{X_i - c_j}{X_i - c_k}\right)^{\frac{1}{m-1}}}. \tag{18}$$

where $i = 1, 2, \dots, n; j = 1, 2, \dots, c; k = 1, 2, \dots, c$.

The proof of Theorem 1 is similar to Thong’s method (Thong and Son 2016), which is omitted here. The pseudocode of the optimized n -Pythagorean fuzzy c -means clustering algorithm is summarized in Algorithm 2.

Remark 2 If the value of n in n -Pythagorean membership degree changes, then n -Pythagorean membership degree also changes as follows:

- (1) If $n = 1$, n -Pythagorean membership reverts to intuitionistic fuzzy c -means clustering algorithm.
- (2) If $n = 2$, n -Pythagorean membership reverts to Pythagorean fuzzy c -means clustering algorithm.
- (3) If $n = 3$, n -Pythagorean membership reverts to Fermatean fuzzy c -means clustering algorithm.

The superiority of the n -PFCM method is presented in Example 1.

Example 1 Exploit enrollment data employed by Song and Chissom (1965a, b) to demonstrate the accuracy improvement in n -PFCM on n -Pythagorean fuzzy time series forecasting. The RMSE obtained by Markov weighted prediction based on FCM on the yearly University of Alabama enrollments data is 1040.1466, and the SMAPE is 2.2369. After

Algorithm 2 Pseudocode for optimized n -Pythagorean fuzzy c -means clustering.

Input:

Initialize the $n, c, m = 2, \alpha = 0.8$ and $\varepsilon = 0.05$.

Initialize the C_j^*, μ_{ij}, v_{ij} according to Eq. (16).

Output:

$C_j^*, \mu_{ij}, v_{ij}, \pi_{ij}$.

1. **While** the termination condition is not satisfied,
2. Calculate the cluster centers C_j^* as Eq. (17).
3. Update the μ_{ij}, v_{ij} , and π_{ij} as Eqs. (18–19, 10).
4. **for** assign the μ_{ij} by Eq. (8).
5. Update the π_{ij} by Eq. (10).
6. Update the v_{ij} by Eq. (11).
7. **end for**
8. **end while**
9. **Return** the C_j^*, μ_{ij}, v_{ij} and π_{ij} .

the same Markov weighted prediction based on n -PFCM, the RMSE is decreased to 688.9613, and the MAPE is also decreased to 1.9005, which strongly proves that the n -PFCM method can get a better domain division effect.

3.2 A novel improved Markov weighted prediction method

Markov weighting algorithm has been one of the many choices of fuzzy time series forecasting methods through the introduction of predecessors. As far as it seems, there is still for improvement in this method in terms of prediction accuracy. This paper further improves the Markov weighted prediction method used by Alyousifi et al. (2020). The parameter λ is introduced to enhance its use of the singular observation information, and an improved Markov weighting algorithm (IMW) is proposed. The specific implementation process is as follows:

The Markov weighting matrix $O = [O_{ij}]_{n \times n}, (i, j = 1, 2, 3, \dots, n)$, since the number of linguistic value is n , O_{ij} refers to the probability of transition from state p_i to state p_j which can be calculated as follows:

$$O_{ij} = \frac{Q_{ij}}{Q_i}, \tag{19}$$

where Q_{ij} and Q_i represent transition time from state p_i to state p_j and the total number of observations in state p_i .

Thus, Markov weighted matrix O can be given as follows:

$$O = \begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1n} \\ O_{21} & O_{22} & \cdots & O_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ O_{n1} & O_{n2} & \cdots & O_{nn} \end{bmatrix}, \tag{20}$$

where $O_{ij} \in [0, 1]$ and $\sum_{j=1}^n O_{ij} = 1$.

The preliminary calculation equation of the predicted value is defined as follows:

$$Forecast(t + 1) = C(t) \cdot O(t). \tag{21}$$

where (\cdot) is the matrix product operator, $Forecast(t + 1)$ is a forecast value at $(t + 1)$, $C(t)$ is a $1 \times n$ matrix, $O(t)$ is an $n \times 1$ matrix. The value of the element c in the C matrix is the midpoints of the cluster of the corresponding observation.

Example 2 Exploit the enrollment data employed by Song and Chissom (1965a, b) for demonstration. The number of enrollment in 1972 is predicted by the 1st-order n -Pythagorean fuzzy time series model. The transition probabilities of 1971 enrollment were assumed to be $Q_{11} = \frac{2}{3}$ and $Q_{12} = \frac{1}{3}$. The cluster centers c_1 and c_2 were 13,524.69, 147,46.23, respectively. Then $Forecast(1972) = C(1971) \cdot O(1971) = c_1 Q_{11} + c_2 Q_{12} = 13,524.69 \times \frac{2}{3} + 14,746.23 \times \frac{1}{3} = 13,931.87$.

Because of the calculation method, it is possible for the predicted values obtained here to have equal results. To solve this problem, further adjustments to the predicted values are needed. Based on the Alyousifi et al. (2020) method, this paper introduces the parameter λ and proposes a new method to improve Markov weighted prediction. The predicted value can be adjusted with Eq. (22):

$$Fore_{new}(t + 1) = Forecast(t + 1) \pm \lambda |(Forecast(t) - C(t))|, \tag{22}$$

where $\lambda \in [0, 1]$ was hired to adjust the impact of the previous prediction on the next period. $Forecast(t)$ is the prediction value obtained by exploiting Eq. (21). $C(t)$ is the cluster centroid value corresponding to the previous period.

Remark 3 If the value of λ in Eq. (22) changes, then the improve Markov weighted prediction also changes as follows:

- (1) If $n = 0$, the improve Markov weighted prediction reverts to the prediction of Lee (Li et al. 2009).
- (2) If $n = 1$, the improve Markov weighted prediction reverts to the Markov weighted prediction of Alyousifi et al. (2020).

Example 3 Continue to exploit the enrollment data employed by Song and Chissom (1965a, b) for demonstration. If the data in 1990, 1991, and 1992 belong to the same n -Pythagorean fuzzy set p_7 , the transition probabilities of p_7 was assumed to be $Q_{77} = \frac{3}{3}$. The cluster centers c_7 were 19,144.31. Then, the predicted value obtained by only employing Eq. (21) is the same, $Forecast(1990) =$

$Forecast(1991) = Forecast(1992) = 19,144.31$. However, when we introduce the parameter λ and set $\lambda = 0.9$, the situation changes.

$$Fore_{new}(1990) = Forecast(1990) + 0.9 \times |18,970 - 19,152.04| = 18,980.474, Fore_{new}(1991) = Forecast(1991) + 0.9 \times |19,328 - 19,152.04| = 19,302.674, Fore_{new}(1992) = Forecast(1992) + 0.9 \times |19,337 - 19,152.04| = 19,310.774.$$

4 n -Pythagorean fuzzy time series and its forecasting method (n -IMWPFM)

4.1 Several definitions of the n -Pythagorean fuzzy time series

In the extant literature, there have been a lot of studies on intuitionistic fuzzy time series and even the generalization of intuitionistic fuzzy sets such as picture fuzzy sets, which have been used to establish time series models. However, n -Pythagorean fuzzy time series has not yet appeared. In this section, the concept and model description of n -Pythagorean fuzzy time series are proposed for the first time.

Definition 9 (*n -Pythagorean Fuzzy Time Series*) Let X_i be a time series with real-world data. p_1, p_2, \dots, p_c are n -Pythagorean fuzzy sets on the universal set. Then, the n -Pythagorean fuzzy time series $P(t)$ can be expressed as:

$$P(t) = \{X_i, \mu_{p_1}(t), \mu_{p_2}(t), \dots, \mu_{p_c}(t), \nu_{p_1}(t), \nu_{p_2}(t), \dots, \nu_{p_c}(t)\}. \tag{23}$$

where $\mu_{p_i}(t)$ and $\nu_{p_i}(t)$ are membership and non-membership of the i th observation in the j th n -Pythagorean fuzzy set $(p_j(t), j = 1, 2, \dots, c)$, c is the cluster of the X_i . The value of the $\mu_{p_i}(t)$ and $\nu_{p_i}(t)$ can be obtained by n -PFCM, respectively.

Remark 4 If the value of n in n -Pythagorean fuzzy time series changes, then n -Pythagorean fuzzy time series also changes as follows:

- (1) If $n = 1$, n -Pythagorean fuzzy time series reverts to intuitionistic fuzzy time series.
- (2) If $n = 2$, n -Pythagorean fuzzy time series reverts to Pythagorean fuzzy time series.
- (3) If $n = 3$, n -Pythagorean fuzzy time series reverts to Fermatean fuzzy time series.

Definition 10 (*First-order n -Pythagorean fuzzy time series forecast model*) Let $P(t)$ be a n -Pythagorean fuzzy time series. $p_j(t) \in P(t)$ is the n -Pythagorean fuzzy set on the universal set. If $p_j(t)$ is only determined by $p_j(t - 1) \in P(t - 1)$, then the first-order n -Pythagorean fuzzy time series

forecast model is $p_i(t - 1) \rightarrow p_j(t)$ or $P(t - 1) \rightarrow P(t)$; it can also be expressed as:

$$P(t) = \text{fun} \{P(t - 1)\}, \tag{24}$$

where $P(t)$ is the set of n -Pythagorean fuzzy time series with the t th time stamp, $t - 1$ is the $(t - 1)$ th time stamp, fun can be any model, such as fuzzy logic model, statistical model, machine learning model, and even deep learning model.

Definition 11 (*Time-variant and time-invariant n -Pythagorean fuzzy time series*) Let $P(t)$ be a Pythagorean fuzzy time series, $R_i(t, t - 1) = R_i = \bigcup_j R_{ij}$ represents the fuzzy relationship between $P(t)$ and $P(t - 1)$. If for any time t , there is always $R_i(t, t - 1) = R_i(t - 1, t - 2)$, then $P(t)$ is called a time-invariant time series. Conversely, if the fuzzy logic relationship $R_i(t, t - 1)$ is independent of time, then at any time t , $R_i(t, t - 1) \neq R_i(t - 1, t - 2)$, and it is called a time-variant time series.

Definition 12 (*q -order n -Pythagorean fuzzy time series forecast model*) Let $P(t)$ be a n -Pythagorean fuzzy time series. $p_j(t) \in P(t)$ is the n -Pythagorean fuzzy set on the universal set. If $p(t)$ is not only determined by $p(t - 1)$, that is, $p_{j-q}, p_{j-q+1}, p_{j-q+2}, \dots, p_{j-1} \rightarrow p_j$, then the q -order forecast model can be defined as below:

$$P(t) = \text{fun} \{P(t - q), P(t - q + 1), P(t - q + 2), \dots, P(t - 1)\}, \tag{25}$$

where $P(t - q)$ have the same meaning as the first-order model and q is the order of the forecast model, fun can be any model, such as fuzzy logic model, statistical model, machine learning model, and even deep learning model.

Remark 5 In this paper, the fuzzy logic relation model is selected as the fun function so the first-order model can be changed into the following form:

$$P(t) = P(t - 1) \circ R_i(t, t - 1), \tag{26}$$

where “ \circ ” represents the synthesis operation, and $R_i(t, t - 1) = R_i = \bigcup_j R_{ij}$ represents the fuzzy relationship between $P(t)$ and $P(t - 1)$. The prediction model of q -order is temporarily out of the scope of this paper.

4.2 A novel n -Pythagorean fuzzy time series forecasting method (n -IMWPFCM)

On the basis of the n -Pythagorean fuzzy time series and prediction model, this paper proposed an improved Markov weighted Pythagorean fuzzy time series forecasting method based on n -Pythagorean fuzzy c-means (n -IMWPFCM). The

input of the model is the lagged variable of the time series, which includes membership and non-membership. n -PFCM is employed to construct the n -Pythagorean fuzzy number and divide the optimal universe. This paper exploits the Markov weighting matrix to achieve defuzzification prediction, and the introduction of λ to improve the method of Alyousifi Alyousifi et al. (2020). The value of λ is calculated through experiments. The process increases the calculation accuracy. First of all, this is a pioneering work on n -Pythagorean fuzzy time series forecasting algorithms. Secondly, this method uses the degree of membership and the degree of non-membership in the process of defuzzification and makes better use of the information contained in the original data. Third, compared with the current popular machine learning methods, the prediction accuracy of this method is comparable, but the interpretability of the proposed method is higher and the meaning of the model is clearer. The specific steps of the proposed algorithm are as follows. At the same time, the framework of the algorithm is shown in Fig. 1.

Algorithm 3 An improved Markov weighted n -Pythagorean fuzzy time series forecasting method based on n -Pythagorean fuzzy c-means (n -IMWPFCM)

Step 1. Define the universe of discourse, U , the maximum and minimum values in the original time series data are used as the boundaries of the universe of discourse, $U = \{D_{min}, D_{max}\}$.

Step 2. Apply n -PFCM to train the original time series data to get the n -Pythagorean fuzzy time series. Based on the original time series, algorithm1 is used to construct the n -Pythagorean fuzzy time series. For example, let $c = 3$, and assume that original time series have N observations. Thus,

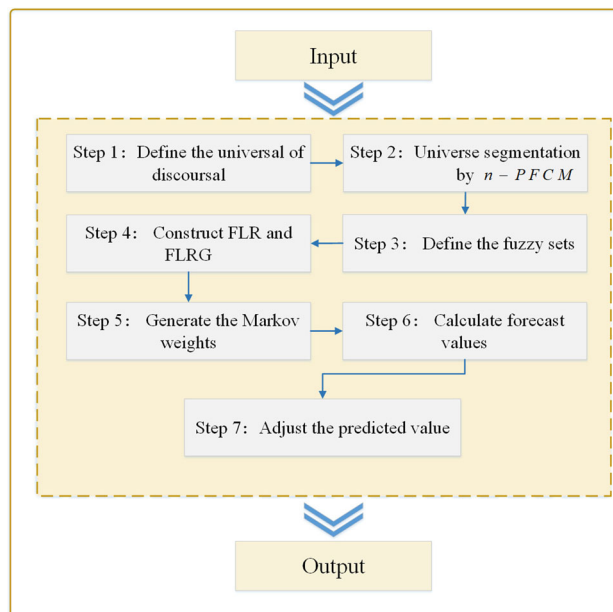


Fig. 1 Framework of the n -Pythagorean time series forecast model

the n -Pythagorean fuzzy time series can be represented as $P(t) = \{X_t, \mu_{p_1}(t), \mu_{p_2}(t), \mu_{p_3}(t), \nu_{p_1}(t), \nu_{p_2}(t), \nu_{p_3}(t)\}$.

Step 3. Define the fuzzy sets p_i on the universal of discourse U by using Eq. (27).

$$p_i = \frac{f_{p_i}(\mu_1)}{\mu_1} + \frac{f_{p_i}(\mu_2)}{\mu_2} + \dots + \frac{f_{p_i}(\mu_n)}{\mu_n}. \tag{27}$$

Step 4. Construct a fuzzy logic relation (FLR) and establish a fuzzy logic relation group (FLRG) to establish a frequency matrix of fuzzy relations between observations. This step can be divided into two substeps. Firstly, the observation results are fuzzy based on the maximum membership value, and the linguistic value of each observation is determined, and the relationship is established in FLR. Such as $p_i \rightarrow p_j$. Secondly, classify all FLRs to establish an FLR group.

Step 5. Generate the Markov weights, that is the transition probability matrix, the matrix of frequencies by the fuzzy relationship groups that have been established in Step 4.

Step 6. Calculate forecast values. The Markov weighted matrix is employed to calculate the predicted value by the product of the defuzzified matrix M and the Markov weighted matrix O , which is given by Eq. (21).

Step 7. Adjust the predicted value to avoid the possibility of duplicate values by using Eq. (22).

Algorithm 3 Pseudocode for n -IMWPFCM.

Input:

X_i ,
 Initialize the $n, c, m = 2, \alpha = 0.8$ and $\varepsilon = 0.05$.
 Initialize the $C_j^*, \mu_{ij}, \nu_{ij}$ according to Eq. (16).

Output:

$Fore_{new}(i + 1)$.
 1. **While** the termination condition is not satisfied,
 2. Calculate the cluster centers C_j^* as Eq. (17).
 3. Update the μ_{ij}, ν_{ij} , and π_{ij} as Eqs. (18–19, 10).
 4. **for** assign the μ_{ij} by Eq. (8).
 5. Update the π_{ij} by Eq. (10).
 6. Update the ν_{ij} by Eq. (11).
 7. **end for**
 8. **end while**
 9. $\mu_i = \operatorname{argmax}(\mu_{ij})$
 10. $Fore_{new}(i + 1) = C_i^* \cdot O(i) \pm \mu_i |X(i) - C_i^*|$.
 11. **Return** the $Fore_{new}(t + 1)$.

5 Numerical example

In this section, the n -IMWPFCM algorithm is applied to two data sets to verify its accuracy and practicality. The first data set is the yearly University of Alabama enrollments data, and the second data set is the U.S. COVID-19 confirmed data,

which comes from Johns Hopkins University (Dong et al. 2020). It is worth noting that in each numerical example, the parameter n in the n -IMWPFCM algorithm is set to 2, which means that the n -IMWPFCM algorithm exploited in the numerical example is IMWPFCM.

All algorithms are implemented in the same environment which is provided by Kaggle Notebooks and Chen’s (Chen and Chung 2006) method is implemented by calling the python package pyFTS (Silva 2018). Both data sets are analyzed by using the classical model which proposed by Chen and Chung (2006), Markov weighted fuzzy time series model based on an optimum partition method proposed by Alyousifi et al. (2020) (MWFCM), Markov weighted fuzzy n -Pythagorean time series model which based on n -Pythagorean fuzzy c -means (n -MWPFCM). Meanwhile, the mean square error (RMSE) root value and the symmetric mean absolute percentage error (SMAPE) value were used to evaluate the validity of the model.

$$MSE = \frac{1}{N} \sum_{t=1}^N (X_t - \hat{X}_t)^2. \tag{28}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (X_t - \hat{X}_t)^2}. \tag{29}$$

$$SMAPE = \frac{100\%}{N} \sum_{t=1}^N \frac{|X_t - \hat{X}_t|}{(|\hat{X}_t| + |X_t|)/2}. \tag{30}$$

5.1 The Yearly University of Alabama enrollments data

The yearly university of Alabama enrollments data is the most commonly employed dataset for fuzzy time series prediction, so the first numerical example in this paper also takes this dataset for prediction and conducts comparative analysis. The dataset is shown in Fig. 2, and descriptive statistics are conducted; the results are shown in Table 1.

Step 1. Define the universe of discourse U from Enrollment data, $U = [13,055, 19,337]$.

Step 2. Apply n -PFCM to train the original time series data to get the n -Pythagorean fuzzy time series. In order to show the clustering result more clearly, it is plotted in Fig. 3.

By clustering this set of data using n -PFCM method, membership and non-membership of each data point to each cluster can be obtained. After repeated experiments, the number of clusters selected in the Yearly University of Alabama enrollments data is 7 ($c = 7$).

Step 3. Define fuzzy sets. For the obtained membership degree, Eq. (27) is exploited to calculate the fuzzy set p_i . The fuzzy set p_i corresponding to each observation is found in Table 2.

Fig. 2 University of Alabama enrollments data

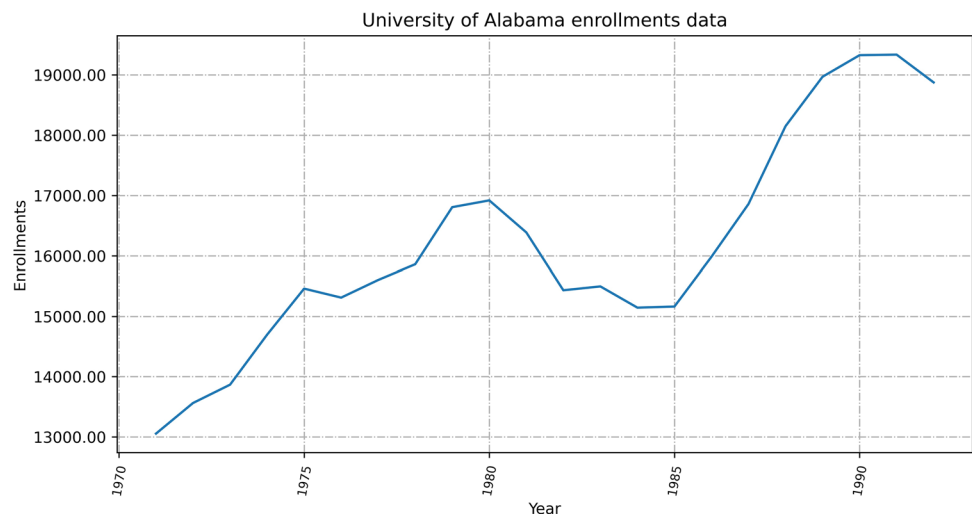


Table 1 Descriptive statistics of University of Alabama enrollment data

Min	1stQu	Median	Mean	3rdQu	Max
13,055	15,200	15,732	16,194	16,904	19,337

Step 4. Construct a fuzzy logic relation (FLR) and establish a fuzzy logic relation group (FLRG). Transform the enrollment data into n -Pythagorean fuzzy numbers and determine the fuzzy logic relationships (FLRs), as observed in Table 2.

Then, determine FLRGs according to the obtained FLRS, and the results are shown in Table 3.

Step 5. Generate the Markov weights based on the matrix of frequencies from Step 4. By defining 7 states for each of the fuzzy sets, matrix $O_{7 \times 7}$ is produced. For example, in the case of FLRG, $p_1 \rightarrow p_1, p_2$. Then, $Q_{11} = 2, Q_{12} = 1,$ and $Q_{13} = 3$. Thus, $O_{11} = \frac{2}{3}$ and $O_{12} = \frac{1}{3}$, otherwise, $O_{1j} = 0$. The results are shown in Table 4.

Step 6. Forecast values are calculated by using Eqs. (21) and (22) based on Markov weights. In addition, the value of λ in this example is set to the degree of membership corresponding to the observation.

The forecasted values are shown in Table 5, the predicted results and the real values are drawn into a line graph (Fig. 4)

for comparison, and the results of the evaluation indicators are plotted as histogram (Fig. 5).

According to the results of the simulation experiment, we can intuitively see that based on the n -Pythagorean fuzzy c-means clustering and then using the Markov weighted method of MWFCM Li et al. (2009) to carry out the de-fuzzed prediction, the prediction accuracy is indeed significantly improved, which demonstrates the effectiveness of the n -Pythagorean fuzzy c-means algorithm. Secondly, the prediction accuracy was improved once again after we introduced λ to improve the Markov weighting method of Alyousifi (n -IMWPFM), whether we used MSE, RMSE and SMAPE to evaluate.

5.2 The new COVID-19 cases data

Coinciding with the outbreak of the COVID-19 epidemic, the global economy is facing tremendous challenges. Although targeted vaccines have been available, the future development trend of the epidemic is still a matter of concern. Predicting the number of new cases is an important basis for judging the development of the epidemic. For this reason, this paper try to establish a n -Pythagorean fuzzy time series prediction model for the new COVID-19 case data in the United States to provide practice for the COVID-19 epidemic prediction

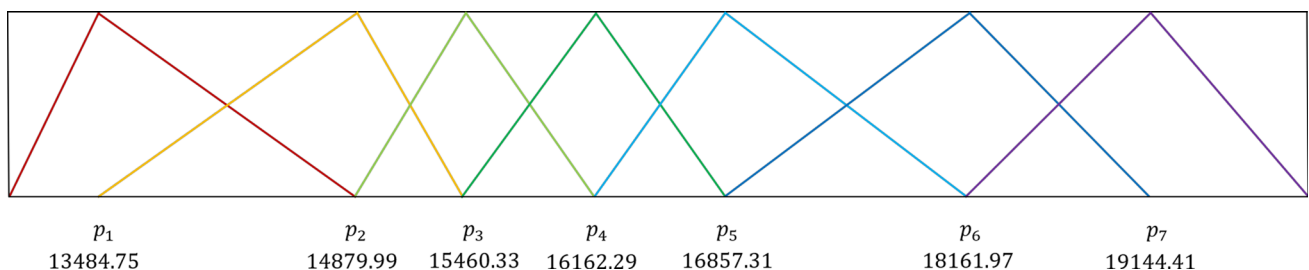


Fig. 3 Results of n -Pythagorean Fuzzy c-means clustering on the yearly of Alabama enrollments data

Table 2 Enrollment data expressed as n -Pythagorean fuzzy numbers

N	Year	Enrollment data	fuzzy sets	FLRs
1	1971	13,055	p_1	–
2	1972	13,563	p_1	$p_1 \rightarrow p_1$
3	1973	13,867	p_1	$p_1 \rightarrow p_1$
4	1974	14,696	p_2	$p_1 \rightarrow p_2$
5	1975	15,460	p_3	$p_2 \rightarrow p_3$
6	1976	15,311	p_3	$p_3 \rightarrow p_3$
7	1977	15,603	p_3	$p_3 \rightarrow p_3$
8	1978	15,861	p_4	$p_3 \rightarrow p_4$
9	1979	16,807	p_5	$p_4 \rightarrow p_5$
10	1980	16,919	p_5	$p_5 \rightarrow p_5$
11	1981	16,388	p_4	$p_5 \rightarrow p_4$
12	1982	15,433	p_3	$p_4 \rightarrow p_3$
13	1983	15,497	p_3	$p_3 \rightarrow p_3$
14	1984	15,145	p_3	$p_3 \rightarrow p_3$
15	1985	15,163	p_3	$p_3 \rightarrow p_3$
16	1986	15,984	p_4	$p_3 \rightarrow p_4$
17	1987	16,859	p_5	$p_4 \rightarrow p_5$
18	1988	18,150	p_6	$p_5 \rightarrow p_6$
19	1989	18,970	p_7	$p_6 \rightarrow p_7$
20	1990	19,328	p_7	$p_7 \rightarrow p_7$
21	1991	19,337	p_7	$p_7 \rightarrow p_7$
22	1992	18,876	p_7	$p_7 \rightarrow p_7$

Table 3 Fuzzy logical relationship groups for the n -PFCM method

Group	Fuzzy logical relationships (FLRs)
G_1	$p_1 \rightarrow p_1(2), p_1 \rightarrow p_2(1)$
G_2	$p_2 \rightarrow p_3(1)$
G_3	$p_3 \rightarrow p_3(5), p_3 \rightarrow p_4(2)$
G_4	$p_4 \rightarrow p_3(1), p_4 \rightarrow p_5(2)$
G_5	$p_5 \rightarrow p_4(1), p_5 \rightarrow p_5(1), p_5 \rightarrow p_6(1)$
G_6	$p_6 \rightarrow p_7(1)$
G_7	$p_7 \rightarrow p_7(3)$

experience. The specific data are obtained by the first-order difference of the USA. COVID-19 confirmed data provided by Johns Hopkins University from January 23, 2020, to April 19, 2021 (Dong et al. 2020). Both the confirmed data and new cases data are drawn as a line chart in Fig. 6, and descriptive statistics are presented in Tables 6 and 7.

Step 1. Define the universe of discourse U from Enrollment data, $U = [0, 300, 310]$.

Step 2. Apply n -PFCM to train the original time series data to get the n -Pythagorean fuzzy time series. In order to show the clustering result more clearly, it is plotted in Fig. 7.

By clustering this set of data using n -PFCM method, membership and non-membership of each data point to each

Table 4 Markov weighted n -PFTS based on the n -PFCM method

Markov weight elements for each group
$p_1 \rightarrow p_1 (\frac{2}{3}), p_1 \rightarrow p_2 (\frac{1}{3})$
$p_2 \rightarrow p_3(1)$
$p_3 \rightarrow p_3 (\frac{5}{7}), p_3 \rightarrow p_4 (\frac{2}{7})$
$p_4 \rightarrow p_3 (\frac{1}{3}), p_4 \rightarrow p_5 (\frac{2}{3})$
$p_5 \rightarrow p_4 (\frac{1}{3}), p_5 \rightarrow p_5 (\frac{1}{3}), p_5 \rightarrow p_6 (\frac{1}{3})$
$p_6 \rightarrow p_7(1)$
$p_7 \rightarrow p_7 (\frac{3}{3})$

cluster can be obtained. After repeated experiments, the number of clusters selected in this paper is 24 ($c = 24$).

Step 3. Define fuzzy sets, construct a fuzzy logic relation (FLR) and establish a fuzzy logic relation group (FLRG). On this basis, the improved Markov-weighted n -PFCM method is used to predict, and the results are plotted in Fig. 8. In addition, the value of λ is set to 0.9 in this example and its value was obtained through experiments.

The prediction results of Fig. 8 can intuitively show that the proposed method can better approximate the true value and therefore has a better prediction effect. Figure 9 shows a histogram drawn from the results of the four forecasting methods under the three indicators of MSE , $RMSE$ and $SMAPE$. From Fig. 9, it can be intuitively seen that the proposed method has the best effect no matter which evaluation indicator is evaluated. This means that, in the face of major public health safety issues, even though the principle of disease transmission is not clear, the use of the n -Pythagorean fuzzy time series forecasting method can provide an effective basis for government decision-making.

6 Conclusion and discussion

The most important contribution of this paper can be summarized as the first time proposals for a fuzzy time series model based on n -Pythagorean fuzzy sets and to construct a prediction method based on this model. The model's inspiration comes from the intuitionistic fuzzy time series and picture fuzzy time series. Using the n -Pythagorean fuzzy set with a wider membership value can be more realistic and effective for the representation of fuzzy information. Secondly, for partitioning the universe of discourse, this paper innovatively proposes a fuzzy c -means clustering method based on n -Pythagorean fuzzy sets (n -PFCM). The proposed method can get the membership degree of real data points more quickly and simply, and the result of clustering is used as the training set of the prediction method. Finally, in the prediction part, this paper improves the Markov weighted prediction method of Alyousifi and introduces λ to make the best use of

Table 5 Comparison of prediction results and accuracy in the Yearly University of Alabama enrollments data (The best results are shown in bold)

Year	Real value	Chen	MWFCM (Alyousifi et al. 2020)	MWPFCM	IMWPFCM
1971	13,055	–	–	–	–
1972	13,563	14,000	14,007.74	13,200.15	13,184.77
1973	13,867	14,000	13,708.14	13,499.75	13,563.74
1974	14,696	14,000	14,438.78	14,438.79	14,110.12
1975	15,460	15,500	16,074.22	16,074.22	15,413.06
1976	15,311	15,500	16,657.01	16,507.23	15,478.39
1977	15,603	16,000	16,656.23	16,508.01	15,358.76
1978	15,861	16,000	16,929.53	16,929.53	15,732.95
1979	16,807	16,000	17,325.37	15,377.77	15,881.65
1980	16,919	17,500	17,000.76	16,945.16	16,819.15
1981	16,388	16,000	17,941.55	16,004.37	16,922.77
1982	15,433	16,000	17,354.46	15,348.68	16,398.89
1983	15,497	16,000	16,630.01	16,534.23	15,438.42
1984	15,145	15,500	16,694.01	16,470.23	15,526.57
1985	15,163	16,000	16,822.23	16,342.01	15,396.79
1986	15,984	16,000	17,369.53	15,794.71	15,421.9
1987	16,859	16,000	17,202.37	17,202.37	17,018.33
1988	18,150	17,500	18,275.42	18,275.42	18,149
1989	18,970	19,000	20,138.62	18,150	18,162.48
1990	19,328	19,000	19,144.31	19,144.31	18,968.64
1991	19,337	19,500	19,511.69	19,144.31	19,315.87
1992	18,876	19,000	19,529.69	19,144.31	18,988.61
MSE	–	195,876.8095	1,081,904.9346	474,667.6705	156,976.8978
RMSE	–	442.5797	1040.1466	688.9613	396.2031
SMAPE	–	2.2369	5.2174	3.4452	1.9005

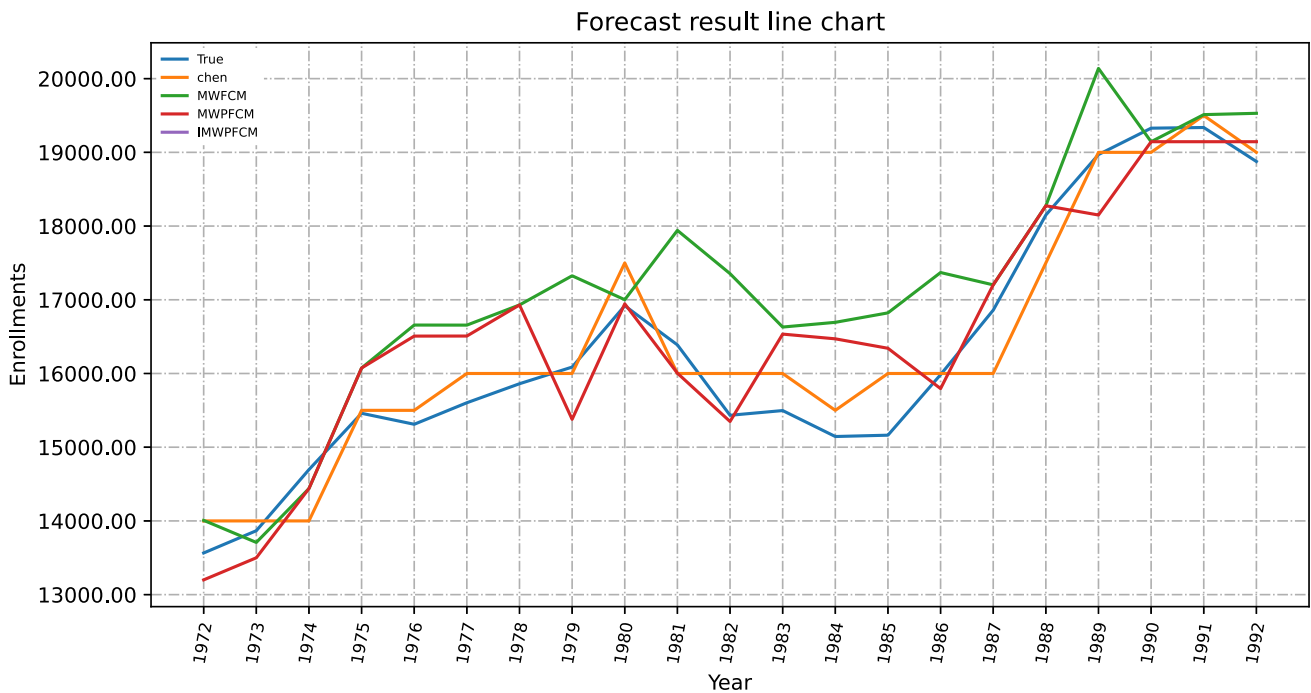


Fig. 4 Forecast result line chart in the Yearly University of Alabama enrollments data

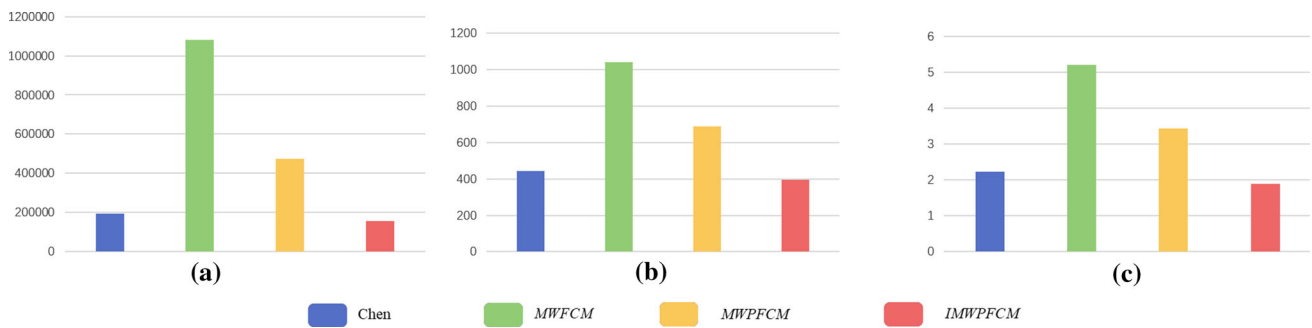


Fig. 5 Forecast result histogram in the Yearly University of Alabama enrollments data. a MSE. b RMSE. c SMAPE

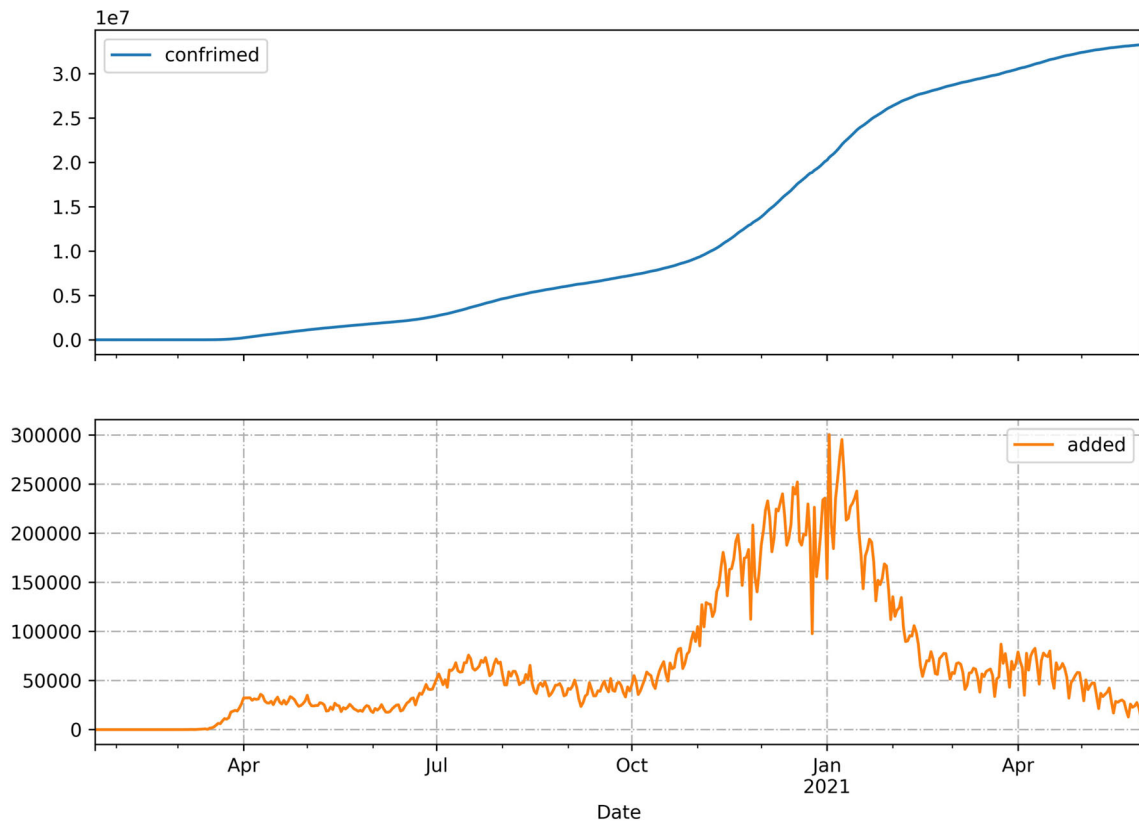


Fig. 6 The US COVID-19 confirmed data and new COVID-19 cases data

Table 6 Descriptive statistics of the new COVID-19 cases data in the USA

Min	1stQu	Median	Mean	3rdQu	Max
0	25,406	49,284	70,062	79,878	300,310

the information contained in the original data, thereby effectively improving the accuracy of model prediction.

The method proposed in this paper is compared with the prediction methods of Chen, MWFCM, MWPFCM. The results clearly demonstrate that the prediction method in this paper effectively improves the prediction accuracy. At the

same time, the advantages of the proposed method in this paper are more noticeable on the small and centralized data set.

In the future research work, the method proposed in this paper can be further improved in the following aspects. First, in terms of partitioning the universe of discourse, the n -Pythagorean fuzzy c -means clustering proposed in this paper can easily obtain the n -Pythagorean membership degree, but there is still a possibility of falling into local optimum, so the clustering method can be optimized for this purpose. Second, in terms of prediction methods, this paper utilizes an improved Markov weighting method, and machine learning or deep learning methods can be used for prediction in the

Table 7 Comparison of prediction results and accuracy in the new COVID-19 cases data (The best results are shown in bold)

Year	Chen	MWFCM (Alyousifi et al. 2020)	MWPFCM	IMWPFCM
MSE	2,516,824,160.2706	1,213,459,996.5654	47,349,215.85609509	41,506,148.9992
RMSE	50,167.9594	34,834.7527	6881.0766	6442.5266
SMAPE	42.3004	38.9377	24.9607	24.8057

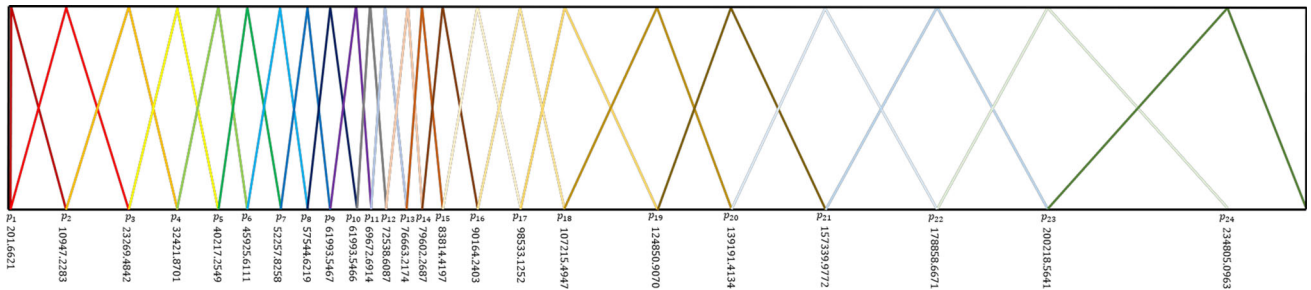


Fig. 7 Results of *n*-Pythagorean Fuzzy *c*-means clustering on the US COVID-19-confirmed data

Fig. 8 Forecast result line chart in the new COVID-19 cases data

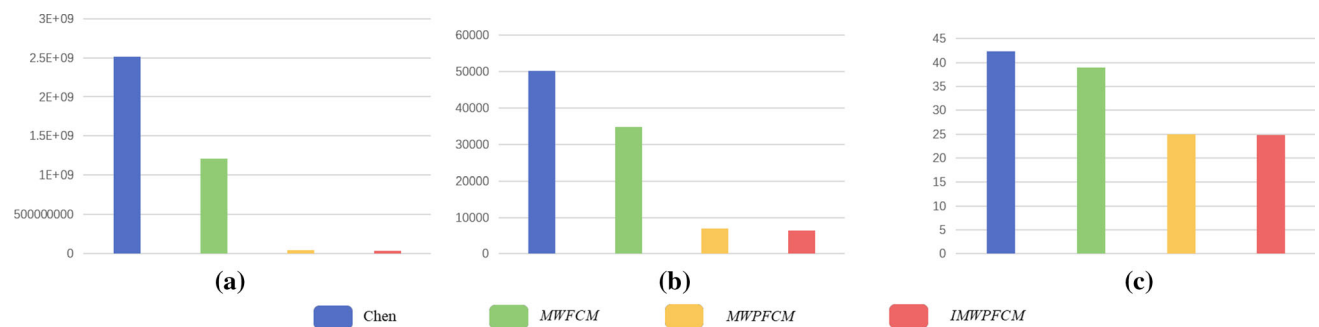
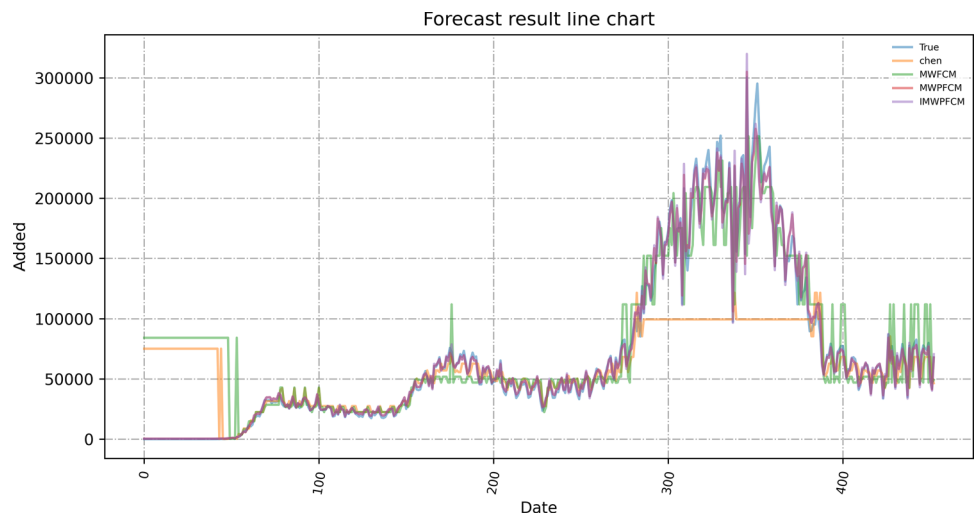


Fig. 9 Forecast result histogram in the new COVID-19 cases data. **a** MSE. **b** RMSE. **c** SMAPE

future. Third, the current model is a first-order univariate model. q -order models for multidimensional variables, such as q -order time series models for spatial data, can be further investigated in the future. Finally, the n -Pythagorean fuzzy time series proposed in this paper has some other parameters, such as n and λ , which are subjectively set. An adaptive method of parameters can be considered for improvement in the future.

Acknowledgements This work was supported by the Graduate Teaching Reform Research Program of Chongqing Municipal Education Commission (No. YJG212022) and Chongqing research and innovation project of graduate students (No. CYS21326).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical statement Articles do not rely on clinical trials.

Human and animal participants All submitted manuscripts containing research which does not involve human participants and/or animal experimentation

References

- Abd Rahman NH, Lee MH, Latif MT (2015) Artificial neural networks and fuzzy time series forecasting: an application to air quality. *Qual Quant* 49(6):2633–2647
- Abhishekh Gautam SS, Singh SR (2018) A score function-based method of forecasting using intuitionistic fuzzy time series. *New Math Nat Comput* 14(01):91–111
- Aladag CH (2013) Using multiplicative neuron model to establish fuzzy logic relationships. *Expert Syst Appl* 40(3):850–853
- Alyousifi Y, Othman M, Faye I (2020) Markov weighted fuzzy time-series model based on an optimum partition method for forecasting air pollution. *Int J Fuzzy Syst* 7:1468–1486
- Atanassov KT (1986) Intuitionistic fuzzy sets. *Fuzzy Sets Syst* 20(1):87–96
- Bas E, Uslu VR, Yolcu U (2014) A modified genetic algorithm for forecasting fuzzy time series. *Appl Intell* 41(2):453–463
- Bas E, Yolcu U, Egrioglu E (2020) Intuitionistic fuzzy time series functions approach for time series forecasting. *Granul Comput* 1–11
- Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c -means clustering algorithm. *Comput Geosci* 10(2–3):191–203
- Bryniarska A (2020) The n -Pythagorean fuzzy sets. *Symmetry* 12(11):1772
- Cai Q, Zhang D, Zheng W (2015) A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression. *Knowl Based Syst* 74:61–68
- Castillo O, Alanis A, Garcia M (2007) An intuitionistic fuzzy system for time series analysis in plant monitoring and diagnosis. *Appl Soft Comput* 7(4):1227–1233
- Chen SM, Chung NY (2006) Forecasting enrollments using high-order fuzzy time series and genetic algorithms. *Int J Intell Syst* 21(5):485–501
- Deng W, Wang G, Zhang X (2015) A novel hybrid water quality time series prediction method based on cloud model and fuzzy forecasting. *Chemom Intell Lab Syst* 149:39–49
- Dincer NG (2018) A new fuzzy time series model based on fuzzy C -regression model. *Int J Fuzzy Syst* 20(6):1872–1887
- Dincer NG, Akkuş Ö (2018) A new fuzzy time series model based on robust clustering for forecasting of air pollution. *Ecol Inform* 43:157–164
- Dong E, Du H, Gardner L (2020) An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 20(5):533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Dunn JC (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *J Cybern* 3(3):32–57
- Duru O, Bulut E (2014) A non-linear clustering method for fuzzy time series: histogram damping partition under the optimized cluster paradox. *Appl Soft Comput* 24:742–748
- Efendi R, Ismail Z, Deris MM (2013) Improved weighted fuzzy time series as used in the exchange rates forecasting of US dollar to ringgit Malaysia. *Int J Comput Intell Appl* 12(01):1350005
- Egrioglu E, Aladag CH, Yolcu U (2013) Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c -means and neural networks. *Expert Syst Appl* 40(3):854–857
- Egrioglu E, Yolcu U, Bas E (2019) Intuitionistic high-order fuzzy time series forecasting method based on pi-sigma artificial neural networks trained by artificial bee colony. *Granul Comput* 4(4):639–654
- Egrioglu E, Bas E, Yolcu U (2020) Picture fuzzy time series: defining, modeling and creating a new forecasting method. *Eng Appl Artif Intell* 88:103367. <https://doi.org/10.1016/j.engappai.2019.103367>
- Fan X, Lei Y, Wang Y (2017) Adaptive partition intuitionistic fuzzy time series forecasting model. *J Syst Eng Electron* 28(3):585–596
- Kumar S, Gangwar SS (2015) Intuitionistic fuzzy time series: an approach for handling nondeterminism in time series forecasting. *IEEE Trans Fuzzy Syst* 24(6):1270–1281
- Li ST, Cheng YC (2010) A stochastic HMM-based forecasting model for fuzzy time series. *IEEE Trans Syst Man Cybern Part B* 40(5):1255–1266
- Li MH, Efendi R, Ismail Z (2009) Modified weighted for enrollment forecasting based on fuzzy time series. *MATEMATIKA Malays J Ind Appl Math* 25:67–78
- Li ST, Kuo SC, Cheng YC et al (2010) Deterministic vector long-term forecasting for fuzzy time series. *Fuzzy Sets Syst* 161(13):1852–1870
- Olej V, Hájek P (2010) IF-inference systems design for prediction of ozone time series: the case of pardubice micro-region. In: International conference on artificial neural networks, pp 1–11
- Ross S (2013) Chapter 12—Markov chain Monte Carlo methods, simulation, 5th edn. Academic Press, Cambridge, pp 271–302. <https://doi.org/10.1016/B978-0-12-415825-2.00012-7>
- Senapati T, Yager RR (2020) Fermatean fuzzy sets. *J Ambient Intell Humaniz Comput* 11:663–674
- Shin Y, Ghosh J (1991) The pi-sigma network: an efficient higher-order neural network for pattern classification and function approximation. In: IJCNN-91-Seattle International Joint Conference on Neural Networks, vol 1. IEEE, pp 13–18. <https://doi.org/10.1109/IJCNN.1991.155142>
- Silva PCL (2018) An open source library for Fuzzy Time Series in Python A open source library for Fuzzy Time Series in Python. *Belo Horizonte*. <https://doi.org/10.5281/zenodo.597359>
- Singh P (2017) High-order fuzzy-neuro-entropy integration-based expert system for time series forecasting. *Neural Comput Appl* 28(12):3851–3868
- Song Q, Chissom BS (1965a) Forecasting enrollments with fuzzy time series part I. *Fuzzy Sets Syst* 54(1):1–9
- Qiang S, Chissom BS (1965b) Forecasting enrollments with fuzzy time series part II. *Fuzzy Sets Syst* 62(1):1–8
- Song Q, Chissom BS (1993) Fuzzy time series and its models. *Fuzzy Sets Syst* 54(3):269–277

- Thong PH, Son LH (2016) Picture fuzzy clustering: a new computational intelligence method. *Soft Comput* 20(9):3549–3562
- Tsaur RC (2012) A fuzzy time series-Markov chain model with an application to forecast the exchange rate between the Taiwan and US dollar. *Int J Innov Comput Inf Control* 8(7B):4931–4942
- Verma H, Agrawal RK, Sharan A (2016) An improved intuitionistic fuzzy c-means clustering algorithm incorporating local information for brain image segmentation. *Appl Soft Comput* 46:543–557
- Vovan T, Ledai N (2019) A new fuzzy time series model based on cluster analysis problem. *Int J Fuzzy Syst* 21(3):852–864
- Xian SD, Xiao Y, Yang ZJ, Li YH (2018a) A new trapezoidal Pythagorean fuzzy linguistic entropic combined ordered weighted averaging operator and its application for enterprise location. *Int J Intell Syst* 33(9):1880–1899
- Xian SD, Yin YB, Fu MQ, Yu FM (2018b) A ranking function based on principal-value Pythagorean fuzzy set in multicriteria decision making. *Int J Intell Syst* 33(8):1717–1730
- Xian SD, Zhang JF, Xiao Y (2018c) A novel fuzzy time series forecasting method based on the improved artificial fish swarm optimization algorithm. *Soft Comput* 22(12):3907–3917
- Xian SD, Li TJ, Cheng Y (2020) A Novel fuzzy time series forecasting model based on the hybrid wolf pack algorithm and ordered weighted averaging aggregation operator. *Int J Fuzzy Syst*. <https://doi.org/10.1007/s40815-020-00906-w>
- Xu Z, Wu J (2010) Intuitionistic fuzzy C-means clustering algorithms. *J Syst Eng Electron* 21(4):580–590
- Yager RR (2013) Pythagorean fuzzy subsets. In: Joint IFSA world congress and nafips annual meeting, pp 57–61
- Yager RR (2014) Pythagorean membership grades in multicriteria decision making. *IEEE Trans Fuzzy Syst* 22(4):958–965
- Yager RR, Abbasov AM (2013) Pythagorean membership grades, complex numbers, and decision making. *Int J Intell Syst* 28(5):436–452
- Zadeh LA (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zhao W, Jiu-Lun F, Hao L (2014) Intuitionistic fuzzy C-means clustering algorithm incorporating local information for image segmentation. *Appl Res Comput* 5358:308–317

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.