A dynamic fuzzy membership degree prediction approach to stock time series

Zhao Chunlan, Cao Xijuan and Xie Xiangjun

July 18, 2019

A dynamic fuzzy membership degree prediction approach to stock time series

1st ChunLan Zhao*
College of Science
Southwest Petroleum University
Chengdu, China
email address:308303451@qq.com

2nd XiJuan Cao*
College of Science
Haojing college of Shaanxi University
of Science & Technology
Xi'an, China
email address:309676730@qq.com

3rd Xiangjun Xie
College of Science
Southwest petroleum university
Chengdu, China
email address:1391005814@qq.com

***Abstract*—Fuzzy time series analysis is the most successful method in enrollment prediction, stock index forecasting, and temperature prediction. It often suffers from high time complexity and low prediction accuracy due to equidistant partition and formulation of fuzzy relationships. In this paper, we propose a concise method called dynamic fuzzy membership degree prediction (DUMP) with four steps for stock time series. First, a number of fuzzy membership degree time series are constructed from the original one. Second, respective prediction models are built with these time series. Third, dynamic fuzzy membership degrees are predicted using these models. Fourth, the final prediction is obtained through the fuzzification of the degree of membership. Comparison study is conducted on 196 stock price time series across one year in comparison with two state-of-the-art approaches. Results show that our approach generally outperforms existing ones in terms of MSE, MAE and MAPE.**

***Index Terms*—dynamic membership function; defuzzification rule; stock price prediction; upper and lower limits.**

I. INTRODUCTION

The prediction of stock market trends is very complex due to the inherent noisy environment and high volatility related to the daily market trends [1]. And stock price movement is nonlinear, complicated, nonparametric and chaotic [2]. Thus, predicting stock price and its trend is one of the most challenging applications of time series analysis. Many investors and professional analysts have made significant contributions to this issue.

Over the past few years, a large body of methods for predicting stock price have been developed. These methods contains many traditional analysis methods, namely logistic regression [3], ARMA model, ARIMA model, and autoregressive moving average with exogenous [4], and many artificial intelligence approaches, namaly artificial neural networks (ANN) [5, 6], support vector machine (SVM) [7–9], and k-nearest neighbors (KNN)[9].

Recently, fuzzy time series method can produce accurate forecasting results due to the handling capability of linguistic value datasets [10, 11]. Song and Chissom presented the concept and model of the fuzzy time series[12, 13]. Then, fuzzy time series model has been improved the from three categories according to the technique used to partition fuzzy intervals. The first category takes equidistant partition. The minimum and maximum values of the sample data were rounded upward and downward to determine the universe classification [12–15]. Then, they took an integer as the length of the interval to uniformly divide the universe based on the size of the universe. The second category considers sample distribution. They adjust the length of linguistic intervals, define a new distance formula and divide intervals, and determine the number of intervals according to the density of the samples, the distance distribution between samples, and the statistical peak of the samples [16, 17]. The third category adopts machine learning algorithms. Optimization algorithms [18, 19] and clustering algorithms [20] are used to find the partition method of the optimal fuzzy subset. In addition, prediction accuracy improvement techniques include weight adjusting [21], hybrid models [22–28], etc. However, the prediction accuracy rates of these methods still exist low prediction accuracy, complex calculation problems.

In this paper, we propose a dynamic fuzzy membership degree prediction (DUMP) with four steps for stock time series. First, six groups of fuzzy membership degree time series are constructed from the original one. We choose the Γ distribution function as membership function, introduce fuzzy technology to define the fuzzy set. Dynamic fuzzy membership degree time series are established. Compared with the traditional fuzzy time series method, there is no need to establish a fuzzy logical relation matrix, which greatly simplifies the calculation process. Second, respective prediction models are built with these time series. Time series analysis can analyze the correlation degree between the sequence data more effectively. We design a fuzzy time series analysis approach based DUMP and TSA. To analyze dynamic fuzzy membership degree properties, we establish time series analysis model to predict the trend of the next moment. Third, dynamic fuzzy membership degrees are predicted using these models. We can predict the membership values for the next four days using the optimal model. Finally, the final prediction is obtained through the fuzzification of the degree of membership. We determine the group number according to prediction results based on time series analysis, revise the membership degree.

This research is supported by the National Science Foundation of China under Grant.

Experiments were performed with 196 datasets to quantify the performance of the DUMP method and compare it with existing methods. We use 192 datasets as training sets, and the rest are as test sets to verify the model. Comparing with two state-of-the-art approaches: ARIMA and Yu's model [17], the maximum MSE, MAE, MAPE of the DUMP method are 0.1923, 0.3098, 0.768%, respectively, which are all less than those of ARIMA and Yu's model. The results show that the DUMP method is more likely to be more accurate than the existing ones in terms of MSE, MAE and MAPE.

This paper is organized as follows. In section II, we briefly provides a overview of the theoretical literature. In Section III, the related definitions and modeling steps of proposed method are described. In Section IV, we extend this model to solve the forecasting problems of stock price. In Section 5, according to the prediction results analysis of various methods, the model is evaluated reasonably. In section 6, it makes a brief summary of the research content.

II. Preliminaries

In this section the necessary concepts of time series analysis, fuzzy time series and K-means method are presented. The first part of this section is about time series analysis approach while the next two sections are fuzzy time series and K-means method respectively.

A. Time series analysis

In 1970, Box and Jenkins published a book called $\ll Time\ Series\ Analysis\ Forecasting\ and\ Control\ \gg$, which systematically expounded the principles and methods of identifying, estimating, testing and predicting of ARIMA [29]. This section briefly introduces the concepts of difference operation, ARMA (p,q) and ARIMA (p,d,q).

Definition 1: Let $X = \{x_1, \ldots, x_n\}$ be a time series and $1 \leq t \leq n$. The first order difference between x_t an x_{t-1} is

$$\nabla x_t = x_t - x_{t-1}, \tag{1}$$

the p oerfer difference of X is

$$\nabla^p x_t = \nabla^{p-1} x_t - \nabla^{p-1} x_{t-1}, \tag{2}$$

and the k step difference is

$$\nabla k = x_t - x_{t-k}. \tag{3}$$

Definition 2: A non-centralized $ARMA(p, q)$ model is

$$x_t = \mu + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t. \tag{4}$$

where $\varepsilon_t \sim$ WN $(0, \sigma_\varepsilon^2)$, $\Phi(B)$ = 1- $\Phi_1 B$ - $\Phi_2 B^2$ - ...- $\Phi_p B^p$, it is called the p order autoregressive coefficient polynomial; $\Theta(B)$ =1- $\theta_1 B$ - $\theta_2 B^2$ - ... - $\theta_q B^q$ is called the q order moving average coefficient polynomial. B is known as delay operator, which is designed to make the model simple.

The idea of the $ARIMA$ model is to transform the non-stationary sequence into a stationary sequence by the difference operation. If the differential sequence is white noise sequence, then the $ARIMA$ model is establishded for the differential sequence.

The $ARIMA(p, d, q)$ model is

$$\nabla^d x_t = \mu_i + \frac{\Theta(B)}{\Phi(B)} \varepsilon_t. \tag{5}$$

where $\nabla^d = (1 - B)^d$; $\mu_i (i = 1, 2, \ldots, n)$ is the mean of the sequence, is the number of sequences; $\Phi(B)$ = 1- $\Phi_1 B$ - $\Phi_2 B^2$ - ...- $\Phi_p B^p$ and $\Theta(B)$ =1- $\theta_1 B$ - $\theta_2 B^2$ - ...- $\theta_q B^q$ are respectively the autoregressive coefficient polynomial and moving average coefficient polynomial of a stationary invertible $ARMA(p, q)$ model.

B. Definitions of fuzzy time series

Few concepts of fuzzy time series were first proposed by Song and Chissom [12]. Song and Chissom [12, 13] also proposed fuzzy time series forecasting methods. In addition, a fuzzy time series takes its values in fuzzy sets [30].

Definition 3: Let $U = \{u_1, u_2, \ldots, u_n\}$, is regarded as the universe of discourse. A fuzzy set can be defined as

$$A = \frac{f_A(u_1)}{u_1} + \frac{f_A(u_2)}{u_2} + \cdots + \frac{f_A(u_n)}{u_n} = \sum_{i=1}^{n} \frac{f_A(u_i)}{u_i} (1 \leq i \leq n). \tag{6}$$

where f_A is the membership function of the fuzzy set A, $f_A : U \to [0, 1]$, and $f_A(u_i)$ represents the grade of the membership of in the fuzzy set A.

Definition 4: Let $Y(t)(t = \ldots, -2, -1, 0, 1, 2, \ldots)$, a subset of R^1, be the universe of discourse on which fuzzy sets $f_i(t)(t = 1, 2, \ldots)$ are defined and $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$. Then $F(t)$ is called a fuzzy time series defined on $Y(t)$.

Definition 5: If $F(t-1) = A_i$ and $F(t) = A_j$, where A_i and A_j are fuzzy sets, then the fuzzy logical relationship between $F(t-1)$ and $F(t)$ is represented by $A_i \to A_j$, where A_i is called the left-hand side (LHS) and A_j is called the right-hand side (RHS) of the fuzzy logical relationship.

III. The proposed method

In this section we present the novel forecasting method, fuzzy time series method based on dynamic membership degree, to predict, its major Related definitions and steps will be discussed in the following subsections.

A. The related definitions of the proposed method

Definition 6: Let $Y(t)(t = \ldots, 0, 1, 2, \ldots)$ be a subset of R^1, $\mu_i(t)(t = 1, 2, \ldots, n)$ is finite continuous nonempty subsets in $Y(t)$. The existence of membership functions f, the $f[\mu_i(t)]$ was established. Then

$$Y(t) \xrightarrow{f[\mu_i(t)]} Y_i^*(t) \tag{7}$$

$Y^*(t)$ is a collection of $Y_1^*(t), Y_2^*(t), \ldots, Y_k^*(t)$. $Y^*(t)$ is called the dynamic membership sequence under $f[\mu(t)]$ action.

Definition 7: Let any fixed dynamic membership sequence $Y^*(t)(t = 1, 2, \ldots, n)$ be a subset of R^1, $f_i(t)$ is defined as a fuzzy set in $Y^*(t)$, and $F(t)$ is a collection of $f_1(t), f_2(t), \ldots$. Then $F(t)$ is called a fuzzy time series defined on $Y^*(t)$.

B. Algorithm

In this section, Algorithm 1 describes the framework of DUMP method construction process. Fig.1 is the frame diagram of the model, where the red region is the difference between the proposed method and the other fuzzy time series [25], and the modeling steps of the proposed method are described in detail below.

Algorithm 1 The framework of DUMP method

Input: Time series $X = \{x_1, \ldots, x_n\}$.
Output: The defuzzification results.
Method: DUMP method.

1: U; // Define the universe of discourse;
2: Partition U into subintervals $U = \{\mu_1, \mu_2, \ldots, \mu_n\}$, define fuzzy sets A_i on U and fuzzy the observed data;
3: $\{Y_1^*(t), Y_2^*(t), Y_3^*(t), \ldots, Y_i^*(t), \ldots, Y_k^*(t)\}$ // DUMP sequence;
4: Set up time series analysis model;
5:Establish defuzzification rules;
(1) Determine the group number;
(2) Revise the membership degree.

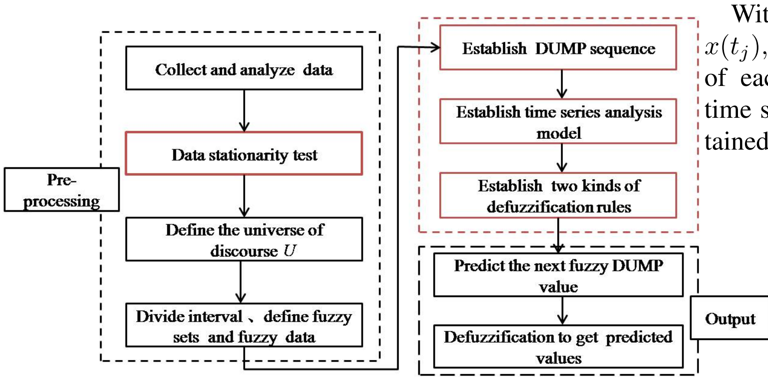

Fig. 1: The structure of the model

Step 1: Define the universe of discourse U.

The U is defined as $U = [D_{min} - \xi_1, D_{max} + \xi_2]$, where D_{min} and D_{max} are respectively the minimum and the maximum of difference sequence, which is obtained by the smooth processing of the original stock price data, and ξ_1 and ξ_2 are two proper positive real values, respectively, to cover the noise of the testing data.

Step 2: Divide the interval, define fuzzy sets and fuzzy the observed data.

The U is divided into k class according to the k-means clustering method, the k clusrer center is sorted from samll to large. Assuming their cluster center is $\nu_1, \nu_2, \nu_3, \ldots, \nu_i, \ldots, \nu_k$, respectively, and the midpoint b_i of two adjacent cluster centers ν_i and ν_{i+1} is taken as the boundary of the sub-domain $\mu_i (1 \le i \le k)$. Then $\mu_1(t) = [D_{min} - \xi_1, \frac{\nu(1)+\nu(2)}{2}]$, $\mu_2(t) = [\frac{\nu(1)+\nu(2)}{2}, \frac{\nu(2)+\nu(3)}{2}]$, $\mu_3(t) = [\frac{\nu(2)+\nu(3)}{2}, \frac{\nu(3)+\nu(3)}{2}]$, …, $\mu_i(t) = [\frac{\nu(i-1)+\nu(i)}{2}, \frac{\nu(i)+\nu(i+1)}{2}]$, …, $\mu_k(t) = [\frac{\nu(k-1)+\nu(k)}{2}, D_{max} + \xi_2]$.

Where $b_1 = \frac{\nu(1)+\nu(2)}{2}$, $b_2 = \frac{\nu(2)+\nu(3)}{2}$, $b_3 = \frac{\nu(3)+\nu(4)}{2}$, …, $b_i = \frac{\nu(i)+\nu(i+1)}{2}$, …, $b_{k-1} = \frac{\nu(k-1)+\nu(k)}{2}$.

According to the actual situation and people understand the ambiguity of the problem, the intervals are interpreted semantically $A_1, A_2, \ldots, A_i, \ldots, A_k$, in a way that natural language can understand, among them, A_i corresponds to the ith fuzzy concepts, that is, fuzzy sets.

Step 3: Establish dynamic fuzzy membership degree sequence.

Existence of membership function f that each sub interval $\mu_i(t)$ can establish the corresponding membership functions under the f action. Shortly, different sub- intervals can correspond to the different membership function and membership function of each sub-interval $\mu_i(t)$ is defined as:

$$\mu_1(t) \xrightarrow{f} f[\mu_1(t)] \tag{8a}$$

$$\vdots \tag{8b}$$

$$\mu_i(t) \xrightarrow{f} f[\mu_i(t)] \tag{8c}$$

$$\vdots \tag{8d}$$

$$\mu_k(t) \xrightarrow{f} f[\mu_k(t)] \tag{8e}$$

With the change of time t, $Y(t) = \{x(t_1), x(t_2), x(t_3), \ldots, x(t_j), \ldots, x(t_n)\}$ is taken into the membership function of each sub-interval defined above, the corresponding new time series $Y_1^*(t), Y_2^*(t), Y_3^*(t), \ldots, Y_i^*(t), \ldots, Y_k^*(t)$ are obtained, which are defined as new time series:

$$Y(t) \xrightarrow{f[\mu_1(t)]} Y_1^*(t) \tag{9a}$$

$$\vdots \tag{9b}$$

$$Y(t) \xrightarrow{f[\mu_i(t)]} Y_i^*(t) \tag{9c}$$

$$\vdots \tag{9d}$$

$$Y(t) \xrightarrow{f[\mu_k(t)]} Y_k^*(t) \tag{9e}$$

As previously known, the original time series is $Y(t)$, then according to the above definition, that is

$$Y_1^*(t) = \{x_1(t_1), x_1(t_2), x_1(t_3), \ldots, x_1(t_j), \ldots, x_1(t_n)\} \tag{10a}$$

$$\vdots \tag{10b}$$

$$Y_i^*(t) = \{x_i(t_1), x_i(t_2), x_i(t_3), \ldots, x_i(t_j), \ldots, x_i(t_n)\} \tag{10c}$$

$$\vdots \tag{10d}$$

$$Y_k^*(t) = \{x_k(t_1), x_k(t_2), x_k(t_3), \ldots, x_k(t_j), \ldots, x_k(t_n)\} \tag{10e}$$

And $Y^*(t) = \{Y_1^*(t), Y_2^*(t), Y_3^*(t), \ldots, Y_i^*(t), \ldots, Y_k^*(t)\}$, then $Y_i^*(t)$ is called the sub-dynamic membership sequence under $f[\mu_i(t)]$ action, and $Y^*t)$ is called the dynamic membership sequence under $f[\mu(t)]$ action.

In this paper, we choose the sharp Γ distribution as the membership function, then the membership function of sub-domain is designed as

$$f[\mu_i(t)] = \begin{cases} e^{p(x-\nu_i)} & x \le \nu_i; \\ e^{-p(x-\nu_i)} & x > \nu_i. \end{cases} \tag{11}$$

where $p > 0$. It can be seen that when the data is closer to the center point, the membership degree is larger, on the contrary, the membership degree is smaller. It meets the basic requirements of clustering, we can determine whether the data belongs to the sub-domain according to the degree of membership of sub-domain.

Step 4: Set up time series analysis model.

Next, we do time series analysis for each sub-domain, and the structure of the time series is in Algorithm 2. We finally get the prediction of fuzzy sets based on the analysis and verification steps.

Algorithm 2 Framework of time series analysis

Input: Membership degree sequence.
Output: Predicted value.
Method: Time series analysis.

1: The sequence of stationary test and white noise inspection;
2: The order recognition of model;
3: Parameter estimation of model;
4: The parameter test and residual test of model;
5: Using the established model to predict the next value.

Step 5: Establish defuzzification rules.

We can predict the membership of each sub-domain in turn by time series modeling. $x_0^*(t_1), x_0^*(t_2), \ldots, x_0^*(t_s)$; $x_1^*(t_1), x_1^*(t_2), \ldots, x_1^*(t_s)$; $\ldots$; $x_k^*(t_1), x_k^*(t_2), \ldots, x_k^*(t_s)$; $x_{k+1}^*(t_1), x_{k+1}^*(t_2), \ldots, x_{k+1}^*(t_s)$, s is the number of test sets. Next,we convert the membership degree of the prediction into the form of fuzzy set,which is called the membership degree vector. Its form is $x_0^*(t_1), x_1^*(t_1), \ldots, x_{k+1}^*(t_1)$; $x_0^*(t_2), x_1^*(t_2), \ldots, x_{k+1}^*(t_2)$; $\ldots$; $x_0^*(t_s), x_1^*(t_s), \ldots, x_{k+1}^*(t_s)$, we can determine whether the data belongs to a group through the analysis and judgment of the data.

(1) Determine the group number

The p of the membership function is relatively small in this research, when the membership degree is generally in the 0.8 to 1, we can determine the data that belongs to a group number. As a matter of fact, the membership degree is generally not between 0.8 and 1. Thereby, we should sort for the membership degree vector, the maximum membership degree is defined as the membership of the group number to which the data belongs. $x_2^*(t_1), x_3^*(t_1), x_1^*(t_1), x_4^*(t_1)$ is taken as a typical example, it indicates that the difference data is in second sub-domain, and is close to the third sub-domain.

(2) Revise the membership degree

When the class of the data is determined, we revise the single digit value to 9 of the maximum membership degree according to the proportion of the maximum membership degree of each group in the membership sequence, which is more in line with the actual situation. Then $x_2^*(t_1), x_3^*(t_1), x_1^*(t_1), x_4^*(t_1)$ is modified to $new x_2^*(t_1), x_3^*(t_1), x_1^*(t_1), x_4^*(t_1)$. We introduce two kinds of defuzzification rules, and the final results are shown in the form of intervals.

The first, we put $new x_2^*(t_1), x_3^*(t_1), x_1^*(t_1), x_4^*(t_1)$ into Eq. (12a, 12b) to get the difference sequence Y.

$$y_{newi}^*(t_j) = -\frac{\log x_i^*(t_j)}{p} + \nu_i \tag{12a}$$

$$y_{newi}^*(t_j) = \frac{\log x_i^*(t_j)}{p} + \nu_i (1 \le j \le s) \tag{12b}$$

After n steps, there are $y_1 = x_{n+1} - x_1$, $y_2 = x_{n+2} - x_2$, $\ldots$, then the predicted values of the stock price are $x_{t+1}^i = y_{t+1-n}^i + x_{t+1-n}^i$, $\ldots$, $x_{t+1+n}^i = y_{t+1}^i + x_{t+1}^i$, and the difference data of smoothing is an interval, the interval of the maximum membership degree is the predicted interval of the stock price.

The second, it is proved that the difference data between the two cluster centers can determine the branch of the membership function according to the membership vector $x_0^*(t_j), x_1^*(t_j), \ldots, x_{k+1}^*(t_j)$. For instance, $x_0^*(t_j), x_{k+1}^*(t_j)$ can determine that the difference data is between the left and right branches of the sub-domain 1 and the left and right branches of the sub-domain k. When a difference data is between ν_i and ν_j, where $i < j$, and put it into Eq. (13a, 13b) to get the difference data. Finally, removing the difference to get stock prices.

$$y_{newi}(t_i) = -\frac{\log x_i^*(t_i)}{p} + \nu_i \tag{13a}$$

$$y_{newj}(t_j) = \frac{\log x_j^*(t_j)}{p} + \nu_j \tag{13b}$$

IV. Empirical applications

In this section, we apply the proposed method to predict the future stock price of Ping An in China. The data are those of China Ping An from 1/2/2014 to 10/4/2014. As shown in the second column of Table 1. Stock price data covering the period from 1/2/2014 to 9/28/2014 are used as the training data set, which illustrate the research procedures in this section, to estimate, and the period from 9/29/2014 to 10/4/2014 is forecast and test the accuracy of the better model for future data. This section focuses on the application of the proposed method in the prediction of stock price.

A. The proposed method

Firstly, Fig.2 is the sequence chart of the original stock price, we can initially determine that the sequence is a non-stationary series. Next, we do the difference processing of the original sequence by the Software programming. The first-order difference is the white noise sequence, which is not of significance. But the second-order difference is non-white noise sequence, so we study the second-order differential sequence.

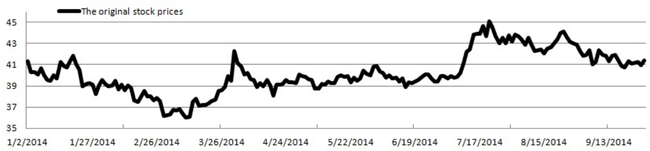

Fig. 2: The original data stock price chart

TABLE I: Fuzzy sets for China Ping An stock price.

Date	Stock price	Fuzzy set	Difference sequences(2)	Fuzzy set
...				
1/20/2014	41.00	A_4	-0.35	A_2
1/23/2014	40.53	A_4	-1.28	A_1
1/24/2014	39.00	A_3	-2.00	A_1
1/25/2014	39.16	A_3	-1.37	A_1
1/26/2014	39.28	A_3	0.28	A_3
...				
6/15/2014	39.70	A_3	-0.04	A_2
6/16/2014	38.87	A_3	-0.55	A_2
6/17/2014	39.34	A_3	-0.36	A_2
6/18/2014	39.28	A_3	0.41	A_3
6/19/2014	39.44	A_3	0.10	A_3
...				
9/28/2014	41.42	A_5	0.15	A_3
9/29/2014	41.14	A_4	0.18	A_3
10/2/2014	41.07	A_4	-0.35	A_2
10/3/2014	40.11	A_4	-1.03	A_1
10/4/2014	40.71	A_4	-0.36	A_2

B. Define U and Establish the sub-domain of difference sequence

We find D_{min} = -2 and D_{max} = 2.39 in two step difference sequence, define $U = [-2.01, 2.4]$. The two step difference sequence is clustered into four classes ,and the final clustering centers are $v_1 = -1.19$, $v_2 = -0.31$, $v_3 = 0.35$, $v_4 = 1.33$ using k-means clustering method that takes into account the distribution and nonuniform problem of the data. Next, $b_1 = -0.75$, $b_2 = 0.02$, $b_3 = 0.84$, the sub-domains are $u_1 = [-2.01, -0.75]$, $u_2 = [-0.75, 0.02]$, $u_3 = [0.02, 0.84]$, $u_4 = [0.84, 2.4]$. Fuzzy set are listed in the last column of Table 1 as an example. The value of difference sequences for 1/20/2014 was -0.35; fuzzify(-0.35) = A_2. The value of difference sequences for 6/19/2014 was 0.10; fuzzify(0.10)= A_3.

C. Establish dynamic membership degree sequence

Establish the membership function for the minimum and maximum values of the two step difference sequence and the cluster centers. The commonly used membership functions include triangle membership function, ladder membership function and Gauss membership function. In this paper, the sharp Γ distribution function is selected as the membership function, the membership function of each sub-domain is defined as in Eq. (14a - 14f).

In particular, when the data belonging to the sub-domain 1 and sub-domain 4, $f[u_0(t)]$ and $f[u_5(t)]$ can determine that the membership function of $f[u_1(t)]$ and $f[u_4(t)]$ should select which branch. Using Software programming and combining the relevant theories in the front and Eq. (14a - 14f), dynamic membership sequence of two step differential sequence are Y_1^*, Y_2^*, Y_3^*, Y_4^*. $0.9 \sim 1$ is the range of membership degree. The number of the third and fourth rows in Table 2 represent the number of belonging to each sub-domain and membership degree $0.9 \sim 1$ in the dynamic membership sequence. Take sub-domain 1 as an example, 26 of the 190 data belong to the sub-domain 1, and the number of membership degree in is only about 14 in Y_1^*. After that, we will revise the stock predicted results according to Step 5 in 3.2.

TABLE II: Belonging to the number of $0.9 \sim 1.0$ in Y_1^*, Y_2^*, Y_3^*, and Y_4^*.

	Sub-domain				Sum2
	1	2	3	4	
Sum1	26	71	71	22	190
$0.9 \sim 1.0$	14	45	47	9	

$$f[\mu_0(t)] = \begin{cases} e^{0.5(x+2)} & x \le -2; \\ e^{-p(x+2)} & x > -2. \end{cases} \tag{14a}$$

$$f[\mu_1(t)] = \begin{cases} e^{0.5(x+1.19)} & x \le -1.19; \\ e^{-0.5(x+1.19)} & x > -1.19. \end{cases} \tag{14b}$$

$$f[\mu_2(t)] = \begin{cases} e^{0.5(x+0.31)} & x \le -0.31; \\ e^{-0.5(x+0.31)} & x > -0.31. \end{cases} \tag{14c}$$

$$f[\mu_3(t)] = \begin{cases} e^{0.5(x-0.35)} & x \le 0.35; \\ e^{-0.5(x-0.35)} & x > 0.35. \end{cases} \tag{14d}$$

$$f[\mu_4(t)] = \begin{cases} e^{0.5(x-1.33)} & x \le 1.33; \\ e^{-0.5(x-1.33)} & x > 1.33. \end{cases} \tag{14e}$$

$$f[\mu_5(t)] = \begin{cases} e^{0.5(x-2.39)} & x \le 2.39; \\ e^{-0.5(x-2.39)} & x > 2.39. \end{cases} \tag{14f}$$

D. Set up time series model

We carry out the time series model of each sub-domain membership sequence using Software programming and combining the modeling process of Algorithm 2, and the sequence diagram of each sub-domain $Y_0^*, Y_1^*, Y_2^*, Y_3^*, Y_4^*, Y_5^*$ are presented in Fig. 3, The blue line represents the original membership degree of each sub-domain. where Y_0^*, Y_5^* represents the membership sequence of the minimum and maximum values of the domain, the sequence always fluctuates between 0 and 1. From these aspects including the sequence diagram, autocorrelation and white noise test, it can be concluded that various membership sequences are stationary and non-white noise sequence, so they have the significance of the research.

We continue to follow the steps in Algorithm 2, the optimal model of each sub-domain is shown in Eqs. (15 - 20), and we map the predicted results of each sub-domain Y_0^*, Y_1^*, Y_2^*, Y_3^*, Y_4^*, Y_5^* in Fig. 3. The red line represents the predicted values of membership degree of each sub-domain.The red line is divided into two parts with the dotted line, one part is used as the training set to fit the original membership degree, the other part is used as the test set to predict the degree of membership at the next moment.

$$Y_0^* = 0.32488 + (1 + 0.67746B^1)\varepsilon_t. \tag{15}$$

$$\begin{aligned} Y_1^* =& 0.571163 + \varepsilon_t/(1 - 0.68837B^1 + 0.63257B^2 - 0.55063B^3 \\ & + 0.30301B^4 - 0.2467B^5 + 0.22106B^6). \end{aligned} \tag{16}$$

$$Y_2^* = 0.755073 + (1 + 0.24884B^1 + 0.14563B^2)\varepsilon_t. \tag{17}$$

$$Y_3^* = (1 + 0.37312B^1 - 0.98504B^2 - 0.38808B^4)\varepsilon_t. \tag{18}$$

$$Y_4^* = 0.53497 + \frac{\varepsilon_t}{1 - 0.51086B^1 + 0.34033B^2 - 0.23784B^3}. \tag{19}$$

$$Y_5^* = 0.392705 + (1 + 0.78809B^1)\varepsilon_t. \tag{20}$$

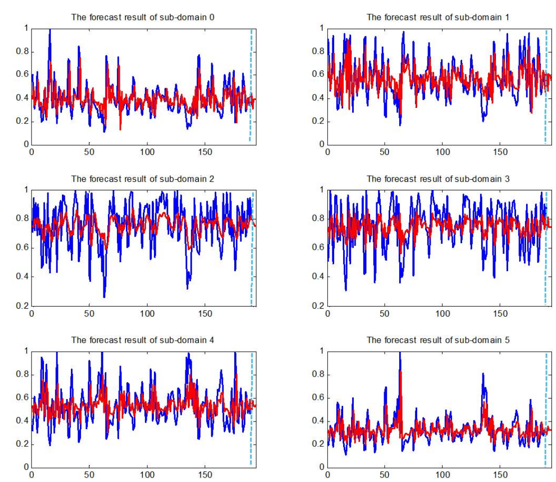

Fig. 3: The forecast result of sub-domain Y_0^*, Y_1^*, Y_2^*, Y_3^*, Y_4^*, Y_5^*

TABLE III: Membership degree of four predicted values in each sub-domain.

Sub-domain	The First	The Second	The Third	The Fourth
Minimum value	0.317919	0.420993	0.415362	0.383467
Sub-domain 1	0.465385	0.608501	0.597549	0.547529
Sub-domain 2	0.778540	0.755023	0.755073	0.755073
Sub-domain 3	0.804314	0.743512	0.736361	0.743222
Sub-domain 4	0.572506	0.530994	0.524769	0.540039
Maximum value	0.355785	0.324880	0.324880	0.324880

The transverse values in Table 3 are the membership degree of the four prediction points belonging to the same sub-domain. For instance, 0.465385 represents that the first prediction point belongs to sub-domain 1 and the membership degree is 0.465385. The longitudinal values are the membership degree of the same prediction point belonging to the different sub-domain. For example, the membership degree of the first prediction point in different sub-domains is 0.465385, 0.778540, 0.804314 and 0.572506.

The degree of membership is more greater, then the probability that the prediction point belongs to the sub-domain is more greater. Therefore, we can judge that the first point belongs to sub-domain 3 and the left branch, the following three points belong to sub-domain 2 and the right branch, and the predicted values of Minimum and Maximum are relatively small, so the four prediction point do not belong to the left branch of sub-domain 1 and the right branch of sub-domain 4.

E. Defuzzification according to debarring rules

In Table 4, we revise membership degree of four predicted values in each sub-domain, which has been marked in Table 4 and the predicted results is done the defuzzification operation to obtain the stock price. More specifically, according to the two kinds methods of defuzzification rules, we can get the predicted results of sub-domain, which is expressed in the form of interval.

TABLE IV: Modified results of predicted membership degree

fuzzy sets	The First	The Second	The Third	The Fourth
A_1	0.465385	0.608501	0.597549	0.547529
A_2	0.778540	**0.955023**	**0.955073**	**0.955073**
A_3	**0.904314**	0.743512	0.736361	0.743222
A_4	0.572506	0.530994	0.524769	0.540039

First, according to the Eqs. (12a, 12b) and Software programming. Table 5 is obtained by the first method of defuzzification rules. Table 5 is divided into two parts, values of the last two lines are the predicted results of defuzzification rule 1, the rest are the predicted results of sub-domain. Through the preceding analysis and Tables 4 and 5, it can be deduced that the first prediction point belongs to the sub-domain 3, the prediction interval is [41.10884, 41.51116], however, the middle gap is relatively large, each sub-domain has a similar value, thereby, we can predict that the result of the first prediction point is [41.10884, 41.15067]. In addition, the second, the third, and the fourth prediction points belong to the sub-domain 2. The results obtained by using defuzzification rule 1 are shown in the last two lines of Table 5. In the same way, according to the Eqs. (13a, 13b) and Software programming. Table 6 represents the predicted results obtained by using defuzzification rule 2.

V. MODEL EVALUATION

This section evaluates the model from two aspects: predicted results and errors. It is described in detail below:

A. Comparison of the results of various methods

The stock price from 9/29/2014 to 10/4/2014 are used for forecasting. The predicted results are summarized in Table 7. The original stock price, the forecasts from Yu's model,

TABLE V: The prediction results of stock sub-domain and defuzzification rule 1.

Date		9/29/2014	10/2/2014	10/3/2014	10/4/2014
The original data		41.14000	41.07000	40.11000	40.71000
Sub-domain 1	Lower bound	38.24022	39.23649	36.02038	36.84181
	Upper bound	41.29978	41.22351	41.13962	41.23819
Sub-domain 2	Lower bound	40.14933	41.01796	39.74740	40.61603
	Upper bound	41.15067	41.20204	40.93260	40.98397
Sub-domain 3	Lower bound	41.10884	41.17726	40.84677	40.93347
	Upper bound	41.51116	42.36274	42.47323	43.30626
Sub-domain 4	Lower bound	41.17454	41.48399	41.21494	41.58176
	Upper bound	43.40546	44.01601	46.02506	46.57824
The prediction	Lower bound	41.10884	41.01796	39.74740	40.61603
Results of stock	Upper bound	41.15067	41.20204	40.93260	40.98397

TABLE VI: The prediction results of stock sub-domain and defuzzification rule 2.

Date		9/29/2014	10/2/2014	10/3/2014	10/4/2014
The original data		41.14000	41.07000	40.11000	40.71000
The prediction	Lower bound	41.10884	41.17726	40.84677	40.93347
Results of stock	Upper bound	41.15067	41.20204	40.93260	40.98397

that is fuzzy time series, the forecasts from ARIMA method and the forecasts using the proposed method are compared in Figs. 4 and 5. In Fig. 4, the prediction of 9/29/2014's is taken as an example. From left to right, the bar chart represents respectively the predicted result of ARIMA, the predicted result of the lower bound, the original stock price, the predicted result of the upper bound and the predicted result of Yu's model. Fig. 5 is similar. Figs. 4 and 5 illustrate intuitively that the proposed method is obviously superior to ARIMA and Yu's model.

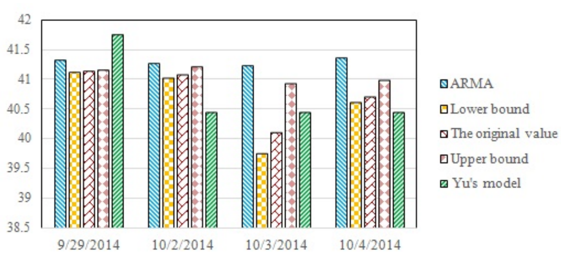

Fig. 4: Comparison of results of the defuzzification rule 1and other methods

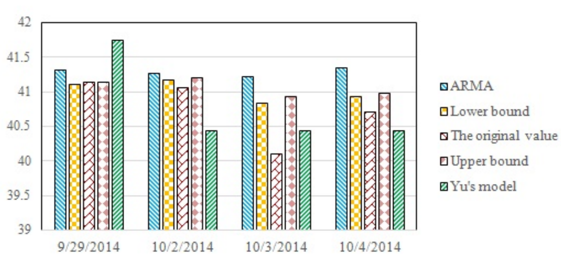

Fig. 5: Comparison of results of the defuzzification rule 2and other methods

B. Comparison of prediction error of various methods

The performance evaluation is assessed by comparing the measuring forecast error criteria via the mean squared error (MSE), mean absolute error (MAE) and the mean absolute percent error (MAPE), and Eqs. (21 - 23) define the MSE, MAE and MAPE.

$MeanSquaredError(MSE)$:

$$MSE = \frac{\sum_{i=1}^{m}(forecast(i) - actual(i))^2}{m} \tag{21}$$

$MeanAbsoluteError(MAE)$:

$$MAE = \frac{\sum_{i=1}^{m}|(forecast(i) - actual(i))|}{m} \tag{22}$$

$MeanAbsolutePercentError(MAPE)$:

$$MAPE = 100\% * \frac{\sum_{i=1}^{m}|(forecast(i) - actual(i))/actual(i)|}{m} \tag{23}$$

where $actual(i)$ and $forecast(i)$ respectively denotes the actual value and the predicted value at time i, m denotes the number of testing period in Eqs. (21 - 23).

The error evaluation results are summarized in Table 7. The performance measures such as mean squared error MSE, mean absolute error MAE and the mean absolute percent error MAPE are considered for evaluating the performance of the proposed method over other prediction methods. In the Table 7, the MSE of the ARIMA and Yu's model prediction respectively are 0.4341 and 0.2349, the MSE of the lower bound and upper bound of the proposed method one are 0.0360 and 0.1923 respectively. Similarly, the MSE of the lower bound and upper bound of the method two are less than that of time series analysis and Yu's model. The MAE and MAPE, of which methods proposed are less than those of the time series analysis and Yu's model. Therefore, the experimental results show that the proposed method is more representative, reliability, scientific, rationality than ARIMA and Yu's model.

TABLE VII: The predicted results and error evaluation of various methods.The original data (Tod), Lower bound of method one (Lbo), Upper bound of method one (Ubo), Lower bound of method two (Lbt), Upper bound of method two (Ubt)

Method	Date				Forecast error criteria via		
	9/29/2014	10/2/2014	10/3/2014	10/4/2014	MSE	MAE	MAPE
Tod	41.14000	41.07000	40.11000	40.71000			
ARIMA	41.32000	41.27000	41.23000	41.35000	0.4341	0.5350	1.322%
Yu's model	41.75000	40.45000	40.45000	40.45000	0.2349	0.4575	1.120%
Lbo	41.10884	41.01796	39.74740	40.61603	0.0360	0.1349	0.334%
Ubo	41.15067	41.20204	40.93260	40.98397	0.1923	0.3098	0.768%
Lbt	41.10884	41.17726	40.84677	40.93347	0.1513	0.2747	0.681%
Ubt	41.15067	41.20204	40.93260	40.98397	0.1923	0.3098	0.768%

VI. Conclusions

In this paper, we have proposed DUMP approach to forecast the stock price of China Ping An. Firstly, the proposed method uses the K -means clustering algorithm to get the cluster center of each cluster and to get the optimal partition of the intervals in the universe of discourse, which can solve the problem of uneven distribution of data, and the midpoint of two adjacent cluster centers is taken as the boundary of the sub-domain. Secondly, we choose the Γ distribution function as the membership function, the membership function of each sub-domain is established to obtain the corresponding dynamic membership sequence, the time series analysis is used to get the predicted value of each sub-domain membership degree. Thirdly, the membership degree of prediction is revised, and then, according to two kinds of the defuzzification rules, we obtain the predicted interval of stock price. From the experimental results shown in Table 7, we can see that the proposed fuzzy forecasting method outperforms the time series method and Yu's model for forecasting the stock price of China Ping An.

ACKNOWLEDGEMENTS

This work is in part supported by National Science Foundation of China under Grant No. 61379089.

REFERENCES

[1] J. L. Ticknor, "A bayesian regularized artificial neural network for stock market forecasting," *Expert Systems with Applications*, vol. 40, no. 14, pp. 5501–5506, 2013.

[2] Y. S. Abu-Mostafa and A. F. Atiya, "Introduction to financial forecasting," *Applied Intelligence*, vol. 6, no. 3, pp. 205–213, 1996.

[3] J. Gong and S. Sun, "A new approach of stock price prediction based on logistic regression model," in *International Conference on New Trends in Information and Service Science*, pp. 1366–1371, 2009.

[4] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, "Time series analysis: Forecasting and control, 5th edition," *Journal of the Operational Research Society*, vol. 22, no. 2, pp. 199–201, 2015.

[5] L. Xi, M. Hou, M. H. Lee, J. Li, D. Wei, H. Hai, and Y. Wu, "A new constructive neural network method for noise processing and its application on stock market prediction ," *Applied Soft Computing Journal*, vol. 15, no. 2, pp. 57–66, 2014.

[6] W. C. Chiang, T. Wu, T. Wu, and R. Wang, "An adaptive stock index trading decision support system," *Expert Systems with Applications An International Journal*, vol. 59, no. C, pp. 195–207, 2016.

[7] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems with Applications An International Journal*, vol. 42, no. 1, pp. 259–268, 2015.

[8] X. Gong, Y. W. Si, S. Fong, and R. P. Biuk-Aghai, "Financial time series pattern matching with extended ucr suite and support vector machine," *Expert Systems with Applications An International Journal*, vol. 55, no. C, pp. 284–296, 2016.

[9] Y. Chen and Y. Hao, "A feature weighted support vector machine and k-nearest neighbor algorithm for stock market indices prediction," *Expert Systems with Applications*, vol. 80, 2017.

[10] M. H. Lee, H. J. Sadaei, and Suhartono, "Introducing polynomial fuzzy time series," *Journal of Intelligent & Fuzzy Systems*, vol. 25, no. 1, pp. 117–128, 2013.

[11] S. H. Javedani and M. H. Lee, "Multilayer stock forecasting model using fuzzy time series.," *The Scientific World Journal,2014,(2014-1-29)*, vol. 2014, no. 2, p. 610594, 2014.

[12] Q. Song and B. S. Chissom, "Fuzzy time series and its model," *Fuzzy Sets & Systems*, vol. 54, no. 3, pp. 269–277, 1993.

[13] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series ł part ii," *Fuzzy Sets & Systems*, vol. 62, no. 1, pp. 1–8, 1994.

[14] S. M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets & Systems*, vol. 81, no. 3, pp. 311–319, 1996.

[15] J. R. Hwang, S. M. Chen, and C. H. Lee, "Handling forecasting problems using fuzzy time series," *Fuzzy Sets & Systems*, vol. 100, no. 1C3, pp. 217–228, 1998.

[16] T. A. Jilani and S. M. A. Burney, "A refined fuzzy time series model for stock market forecasting," *Physica A Statistical Mechanics & Its Applications*, vol. 387, no. 12, pp. 2857–2862, 2008.

[17] H. K. Yu, "A refined fuzzy time-series model for forecasting," *Physica A Statistical Mechanics & Its Applications*, vol. 346, no. 3–4, pp. 657–681, 2005.

[18] C. H. Aladag, U. Yolcu, and E. Egrioglu, "A high order fuzzy time series forecasting model based on adaptive expectation and artificial neural networks," *Mathematics & Computers in Simulation*, vol. 81, no. 4, pp. 875–882, 2010.

[19] E. Egrioglu, C. H. Aladag, U. Yolcu, V. R. Uslu, and N. A. Erilli, "Fuzzy time series forecasting method based on gustafson-kessel fuzzy clustering," *Expert Systems with Applications An International Journal*, vol. 38, no. 8, pp. 10355–10357, 2011.

[20] S. M. Chen and K. Tanuwijaya, "Multivariate fuzzy forecasting based on fuzzy time series and automatic clustering techniques," *Expert Systems with Applications*, vol. 38, no. 8, pp. 10594–10605, 2011.

[21] C. H. Cheng, J. W. Wang, and C. H. Li, "Forecasting the number of outpatient visits using a new fuzzy time series based on weighted-transitional matrix," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2568–2575, 2008.

[22] S. C. Kuo, S. C. Kuo, and C. J. Lin, *An efficient forecasting model based on an improved fuzzy time series and a modified group search optimizer*. Kluwer Academic Publishers, 2017.

[23] M. Y. Chen and B. T. Chen, *A hybrid fuzzy time series model based on granular computing for stock price forecasting*. Elsevier Science Inc., 2015.

[24] F. Ye, L. Zhang, D. Zhang, H. Fujita, and Z. Gong, "A novel forecasting method based on multi-order fuzzy time series and technical analysis," *Information Sciences*, vol. 367–368, pp. 41–57, 2016.

[25] B. Q. Sun, H. Guo, H. R. Karimi, Y. Ge, and S. Xiong, "Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series," *Neurocomputing*, vol. 151, no. 3, pp. 1528–1536, 2015.

[26] F. M. Talarposhti, H. J. Sadaei, R. Enayatifar, F. G. Guimar?es, M. Mahmud, and T. Eslami, "Stock market forecasting by using a hybrid model of exponential fuzzy time series," *International Journal of Approximate Reasoning*, vol. 70, no. C, pp. 79–98, 2016.

[27] S. H. Cheng, S. M. Chen, and W. S. Jian, "Fuzzy time series forecasting based on fuzzy logical relationships and similarity measures," *Information Sciences*, vol. 327, no. C, pp. 272–287, 2016.

[28] C.L. Zhao, X.J. Cao, X.J. Xie, X.H. Yuan and B.Wang, "Environmental Aspects of Drilling Operation Risks: A Probability Prediction based on Fuzzy Time Series," Ekoloji, vol. 27, no.106, pp. 1493–1502, 2018.

[29] G. E. P. Box and G. M. Jenkins, *Time series analysis forecating and control*. HOlden-Day, 1970.

[30] L. A. Zadeh, "Fuzzy sets," in *Fuzzy Sets, Fuzzy Logic, & Fuzzy Systems*, pp. 394–432, 1996.