Neurocomputing 151 (2015) 1528–1536

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Prediction of stock index futures prices based on fuzzy sets and multivariate fuzzy time series

BaiQing Sun [a], Haifeng Guo [a,*], Hamid Reza Karimi [b], Yuanjing Ge [a], Shan Xiong [a]

[a] School of Management, Harbin Institute of Technology, China
[b] Department of Engineering, Faculty of Engineering and Science, University of Agder, 4898 Grimstad, Norway

A R T I C L E I N F O

Article history:
Received 5 May 2014
Received in revised form
15 August 2014
Accepted 11 September 2014
Communicated by Yang Tang
Available online 7 October 2014

Keywords:
Fuzzy set theory
Multivariate fuzzy time series
Stock index futures

A B S T R A C T

This paper makes a prediction of Chinese stock index (CSI) future prices using fuzzy sets and multivariate fuzzy time series method. We select Chinese CSI 300 index futures as the research object. The fuzzy time series model combines the fuzzy theory and the time series theory, thus this model can solve the fuzzy data in stock index futures prices. This paper establishes a multivariate model and improves the accuracy of computation. By combing traditional fuzzy time series models and rough set method, we use fuzzy c-mean algorithm to make the data into discrete. Further more, we deal with the rules in mature modules of the rough set and then refine the rules using data mining algorithms. Finally, we use the CSI 300 index futures to test our model and make a prediction of the prices.

1. Introduction

After the launch of the CSI 300 index futures in China, the study of index futures price variation in the Chinese market is of great significance. In order to solve the problem of financial time series analysis and prediction, the researchers used the exponential smoothing method, ARIMA model and state space model and Kalman filter model. However these models hardly fit nonlinear data. To solve this problem, some scholars propose neural network model which simulates the working principle of the neuron and support vector machine model based on the statistics. Although all of the above models can better solve the precise data analysis prediction, they barely deal with fuzzy data.

With the further research and application of the uncertainty phenomenon of systems, there raise some studies related in this area. In order to solve the fuzzy data, Zadeh proposed the concept of fuzzy sets [1] in 1965 and then Pawlak presented the rough set theory [2]. Song and Chissom put forward the concept of the fuzzy time series, which refers to the time-series data consisting of fuzzy set values [3]. Then there are more researchers studying the fuzzy time series. These studies include how to improve fuzzy time-series models and how to use these models into practice activities. Jilani et al. and Nan et al. combine Heuristic model with the fuzzy time series and integrate detailed questions into the model, which improves the prediction accuracy of the model [4,5]. Further, Chen and Hwang extended the fuzzy time series into Binary model [6]. On the basis of the original model, Huarng and Yu extended the order of the model to *N*-order, and then used the new model to predict the weighted stock index of the Taiwan Stock Exchange and found that the method is useful to predict prices [7]. Meanwhile, Sadaei et al. proposed the fuzzy time series model with the fixed weights [8]. On the basis of Yu's model, Cheng et al. proposed the fuzzy time-series model based on the trend weights [9]. In addition, some researchers utilized the superiority of the neural network in predicting non-linear data and thus combined the neural network model with the fuzzy time-series model [10,11]. These methods are used into many aspects, such as predicting enrollment [12,13], the predicting stock index [14], temperature forecasting [6] and the travel demand forecasting [15], etc. For other studies, fuzzy is used in many fields for improving the accuracy or applicability of the method [16]. For instance, the fuzzy entropy is used to simplify the process of finding optimal segmentation thresholds in order to create a multi-level thresholding method. A new linguistic truth-valued intuitionistic fuzzy lattice is proposed on the basis of intuitionistic fuzzy set [17] and the Takagi–Sugeno (T–S) fuzzy system is adopted to model linear and nonlinear systems [18–20].

This paper can contribute in predicting of stock index futures prices using fuzzy financial data and providing some suggestion for investment decision. In order to solve these problems, this paper combines fuzzy mathematics and statistics, then optimizes fuzzy time series forecasting model and establishes prediction

* Correspondence to: No. 507, School of Management, Harbin Institute of Technology, Harbin 150001, China. Tel.: +86 451 86414019.
E-mail addresses: baiqingsun@hit.edu.cn (B. Sun), haifengguo@hit.edu.cn (H. Guo), hamid.r.karimi@uia.no (H. Reza Karimi), geyuanjing527@gmail.com (Y. Ge), xiongshanxiongshan@126.com (S. Xiong).

http://dx.doi.org/10.1016/j.neucom.2014.09.018
0925-2312/

model based on rough sets and multiple fuzzy time series. At the same time it expends univariate model into multivariate model which can improve the universality of the model and the accuracy on the principle of univariate fuzzy time series analysis. In the last part this paper makes an empirical analysis using the CSI 300 index futures.

The rest of this paper is structured as follows. Section 2 presents theoretical and modeling framework. Section 3 establishes a model based on the rough set and the multivariate fuzzy time series. Section 4 makes a prediction using CSI 300 index futures prices. The conclusion is made in Section 5

2. Theoretical and modeling framework

2.1. Fuzzy time series

The concept of fuzzy time series is put forward by Song and Chissom when they were in the study of how to predict college enrollments [12,13]. This paper will make a brief introduction to fuzzy time series.

Definition 1. Suppose the discourse domain is U and $U=\{u_1, u_2, ..., u_n\}$. The fuzzy set A in U is defined

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \cdots + f_A(u_n)/u_n \tag{1}$$

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

Definition 2. Assume $Y(t)\ (t=\cdots,0,1,2,\cdots)$, a subset of R^1 is the discourse domain of the fuzzy set $f_i(t)$. If $F(t)$ is the collection of $f_i(t)\ (i=1,2,\cdots)$, then $F(t)$ is called the fuzzy time series of the time series $Y(t)\ (t=\cdots,0,1,2,\cdots)$.

Definition 3. If $F(t)$ is only determined by $F(t-1)$, such as $F(t-1) \to F(t)$, then this relation can be presented as $F(t) = F(t-1) \circ R(t,t-1)$, in which $R(t,t-1)$ is the fuzzy relation between $F(t-1)$ and $F(t)$. "$\circ$" is a composition operator, but in the real process, most literatures adopt the max–min operator.

Definition 4. If $F(t-1)=A_i$ and $F(t)=A_j$, then the fuzzy logical relation between the two consecutive observations – $F(t-1)$ and $F(t)$ can be presented as $A_i \to A_j$, in which A_i is called the left-hand side (LHS) and A_j the right-hand side (RHS) of the LHS. The fuzzy logical relations with the same LHS can constitute the fuzzy logical relation group: $A_i \to A_{j_1}, A_i \to A_{j_2}, \cdots$, and the corresponding fuzzy relation group is

$$A_i \to A_{j_1}, A_{j_2}, \cdots \tag{2}$$

Definition 5. If $R(t,t-1)$ has nothing to do with time, that is, $R(t,t-1) = R(t-1,t-2) = \cdots$, then $F(t)$ is considered as a time-invariant fuzzy time series. Otherwise, $F(t)$is time-variant fuzzy time series.

Definition 6. *N*th-order fuzzy logical relationship: If $F(t)$ is related to $F(t-1)$, $F(t-2)$, …, and $F(t-n)$, namely $F(t-n), \cdots, F(t-2), F(t-1) \to F(t)$, then $F(t)$ is called the *n*th-order fuzzy time series.

Definition 7. Suppose there are n fuzzy time series which are $F_1, F_2, \cdots, F_n$, and the linguistic variables of their previous issue are $F_1(t-1) = A_{1_{i_1}}, F_2(t-1) = A_{2_{i_2}}, \cdots, F_n(t-1) = A_{n_{i_n}}$, and $F_1(t) = A_{1_{j_1}}$, that is, $A_{1_{i_1}}, A_{2_{i_2}}, \cdots, A_{n_{i_n}} \to A_{1_{j_1}}$, then this model is named as the *n*-nary fuzzy time series model.

2.2. Fuzzy C-mean clustering algorithm

The fuzzy c-mean algorithm (FCMA) or (FCM) is a clustering method which uses membership to determine the clustering extent. It makes full use of the fuzzy theory and can objectively reflect the real world. At the same time, it has many advantages, such as, processing large data sets, which enables it to be widely applied.

The FCM algorithm is relatively simple, whose main goal is to divide n vectors $x_i(i=1,2,\cdots n)$ into c fuzzy groups which satisfy Eqs. (3)–(5).

$$u_{ij} \in [0,1];\ i=1,2,\cdots c;\ j=1,2\cdots,n \tag{3}$$

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j=1,...,n \tag{4}$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \tag{5}$$

Then, the objective function of FCM is

$$J(U, c_1, ..., c_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m ||x_j - c_i||^2, \tag{6}$$

c_i is the clustering center of the fuzzy group i $||x_j-c_i||^2$ is the Euclidean distance between the no. i clustering center and the no. j data point, and $m \in [1,\infty)$ is a weighted index.

Constructing the following new objective function can make Eq. (6) achieve the necessary minimum condition.

$$\bar{J}(U, c_1, ..., c_c, \lambda_1, ..., \lambda_n) = J(U, c_1, ..., c_c) + \textstyle\sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} u_{ij} - 1)$$
$$= \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^m ||x_j - c_i||^2 + \sum_{j=1}^{n} \lambda_j (\sum_{i=1}^{c} u_{ij} - 1) \tag{7}$$

In Eq. (7), $\lambda_j, j=1,2,\cdots,n$ means the n constrained Lagrange multipliers of Eq. (4). We take the derivative of all input parameters and the necessary conditions that make Eq. (6) to achieve the minimum are:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m}, i=1,2,\cdots,c \tag{8}$$

and

$$u_{ij} = \frac{(1/||x_j - c_i||^2)^{1/(m-1)}}{\sum_{k=1}^{c} (1/||x_j - c_i||^2)^{1/(m-1)}}, i=1,2,\cdots c, j=1,2,\cdots,n \tag{9}$$

From the above-mentioned two necessary conditions, it can be deduced that the fuzzy c-means algorithm is a simple iterative process. In the process of the batch mode, FCM uses the following steps to determine the clustering center c_i:

Step1: Using a random number in [0,1] to initialize the membership matrix U and to let it satisfy the constraint conditions of the Eq. (4).
Step 2: Using Eq. (8) to calculate the c clustering centers.
Step 3: According to Eq. (6), we calculate the value function. If it is less than some definite threshold value or its change which is compared to the value of the previous value function is less than some definite threshold value, then the algorithm stops.
Step 4: Using the Eq. (9) to calculate the new matrix U and returning back to the step 2.

2.3. The concept of rough set

Definition 8. Suppose that $S=(U,R,V,f)$ is a knowledge representation system, in which U stands for the object of study, that is, the domain of discourse, R is the attribute set of objects that can be divided into two disjoint subsets which are the condition attribute C and the decision attribute D and $R=C \cup D$, V is the set of the