Knowledge-Based Systems 74 (2015) 61–68

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A new fuzzy time series forecasting model combined with ant colony optimization and auto-regression

Qisen Cai [a], Defu Zhang [a,*], Wei Zheng [a,*], Stephen C.H. Leung [b]

[a] Department of Computer Science, Xiamen University, Xiamen 361005, China
[b] Department of Management Sciences, City University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 3 May 2013
Received in revised form 20 October 2014
Accepted 6 November 2014
Available online 15 November 2014

Keywords:
Fuzzy time series
Ant colony
Auto-regression
Stock forecasting
Levenberg–Marquardt algorithm

ABSTRACT

This paper presents a new fuzzy time series model combined with ant colony optimization (ACO) and auto-regression. The ACO is adopted to obtain a suitable partition of the universe of discourse to promote the forecasting performance. Furthermore, the auto-regression method is adopted instead of the traditional high-order method to make better use of historical information, which is proved to be more practical. To calculate coefficients of different orders, autocorrelation is used to calculate the initial values and then the Levenberg–Marquardt (LM) algorithm is employed to optimize these coefficients. Actual trading data of Taiwan capitalization weighted stock index is used as benchmark data. Computational results show that the proposed model outperforms other existing models.

1. Introduction

Relationships between sequential set of past data measured over time to forecast future values, are investigated by time series forecasting. Many statistical tools such as regression analysis, moving average, exponential moving average and autoregressive moving average have been used in traditional forecasting. However, these analysis models highly rely on historical data, and the data are required to follow normal distribution to get a relatively good forecasting performance. Moreover, traditional crisp time-series forecasting methods are normally not applicable when the historical data are represented by linguistic values. In order to deal with these kinds of problems, the fuzzy time-series approach has been developed as an alternative forecasting method. It has been proved that the fuzzy time-series can be appropriately applied to datasets of linguistic values to generate forecasting rules with high accuracy.

Past decades have witnessed the development of fuzzy time series approach, since it was first introduced by Song and Chissom [1] in 1993. Different fuzzy time-series models have been applied to solve problems arising in various domains. For example, Song and Chissom [1,2] and Chen [3] developed fuzzy time series models to forecast enrollments of the University of Alabama. To enhance the accuracy of forecast values, Chen [4] proposed high-order fuzzy time series models for forecasting enrollments. Song [5] introduced an autocorrelation function to measure the dependency between the fuzzy data for selecting suitable order for fuzzy time series. Yu [6] proposed a weighted fuzzy time series model to forecast Taiwan capitalization weighted stock index (TAIEX). Own and Yu [7] presented a heuristic higher order model by introducing a heuristic function to incorporate the heuristic knowledge so as to improve TAIEX forecasting. Chen and Chang [8] presented a method for multivariable fuzzy forecasting based on fuzzy clustering and fuzzy rule interpolation techniques. Huarng [9,10] pointed out that the length of intervals affects forecast accuracy in fuzzy time series and proposed a method with distribution-based length and average-based length to reconcile this issue. Keles et al. [11] proposed a model for forecasting the domestic debt by adaptive neuro-fuzzy inference system. Chen et al. [12] presented a new method to forecast TAIEX using fuzzy time series and automatically generated weights of multiple factors. Gangwar and Kumar [13] proposed a computational method of forecasting based on multiple partitioning and higher order fuzzy time series. Chen and Tanuwijaya [14] adopted an automatic clustering technique to overcome the drawback of partition of the universe of discourse. Chen and Kao [15] proposed a new method for forecasting the TAIEX, based on fuzzy time series, particle swarm optimization techniques and support vector machines. Pritpal and Bhogeswar [16] presented a new model based on hybridization of fuzzy time series theory with artificial neural network (ANN). Cheng and Li [17] proposed an enhanced HMM-based forecasting model by developing a novel fuzzy smoothing method to overcome the

* Corresponding authors.
E-mail addresses: dfzhang@xmu.edu.cn (D. Zhang), zhengw@xmu.edu.cn (W. Zheng).

http://dx.doi.org/10.1016/j.knosys.2014.11.003
0950-7051/

problem of rule redundancy and achieve better results. Recently, Wei et al. [18] developed A hybrid ANFIS based on *n*-period moving average model to forecast TAIEX stock. Chen and Chen [19] presented a hybrid fuzzy time series model based on granular computing for stock price forecasting.

While adopting the fuzzy time series model, a reasonable partition of the universe of discourse can significantly enhance the accuracy of the forecasts. Cai et al. [20] presented a genetic algorithm to partition the universe of discourse. According to our review of literature, models [6–15] proposed in recent years also attached great importance to this issue. From another perspective, high-order fuzzy logical relationship methods [4,7] and multiple-factor forecasting methods [8,12,13] are widely employed in hybrid models. However, these two methods have their own drawbacks. On one hand, the high-order method, according to the shape proposed by Chen [4], cannot achieve the expected coverage of the high-order fuzzy logical relationships when the training data do not reach the desired size. On the other hand, the multiple-factor forecasting method is highly limited by the selection of the secondary factor. In addition, such a multiple-factor method also introduces extra noise when more factors are considered. To the best of our knowledge, compared with several univariate models [6,21], the multiple-factor forecasting method can hardly achieve significant improvement in accuracy of forecasting results.

Taking into account the discussion above, this paper focuses on two issues: (1) searching a suitable partition of universe of discourse and (2) taking better advantage of the high-order data. We propose a hybrid model to address these two issues in order to improve forecasting accuracy. For partitioning the universe of discourse, although many statistical methods have been used to achieve efficient results, we believe there still remains a plenty of space for improvement by adopting the Meta-heuristic optimization algorithms, such as the ant colony optimization (ACO) algorithm. ACO is a well-known probabilistic technique for solving computational problems [22,23]. Since finding partition can be defined as a graph search problem, and ACO is commonly thought to be good at solving this type of problem, we adopt ACO in this paper, where the ants are employed to search boundary of each interval and finally obtain a better partition in the research. In order to enhance the application of the historical high-order data, we propose a fuzzy time series model combined with the auto-regression method. The proposed model uses the percentage change as the universe of discourse, the same as that proposed by Stevenson and Porter in [24]. To verify the performance of the proposed model, the TAIEX is used as the experimental dataset and several fuzzy time series models [3,6,8,12,25–27] are used as competitors. The experimental results show that the proposed model gets higher average forecasting accuracy rates than other existing models.

The rest of this paper is organized as follows. Section 2 briefly reviews the definitions of fuzzy time series, ACO algorithm and Levenberg–Marquardt (LM) algorithm. Section 3 presents a novel high-order fuzzy time series model which adopts the classic concept of the auto-regression model. Section 4 presents how ACO is employed to search the suitable partition of fuzzy time series model. Section 5 compares and discusses the forecast results of the proposed model with those of existing models and shows the details of forecast results. The conclusion is provided in Section 6.

2. Brief review of basic concepts

As shown in Fig. 1, the proposed model employs ACO algorithm to search the best partition for high-order fuzzy time series model. Then the LM algorithm is adopted to optimize the coefficients of the fuzzy time series model. In this section, the underlying concepts of fuzzy time series, ACO, and the LM algorithm are introduced.

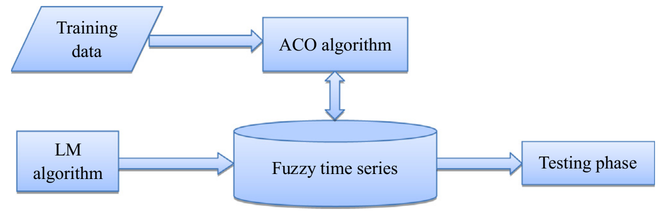

Fig. 1. Structure of the proposed model.

2.1. Fuzzy time series

In the past few decades, research on time series has made progress in dealing with precise figures. However, in real world, people tend to encounter a lot of random fuzzy sequences containing noise. Prediction based on traditional time series appears to be powerless in such situations. Fortunately, it was realized that fuzzy mathematics has a great advantage in solving such problems. As a result, Song and Chissom [1,2] introduced the concept of fuzzy mathematics into time series and proposed the concept of fuzzy time series. We briefly introduce the concepts and notations related to fuzzy time series as follows.

Let U be the universe of discourse, where $U = \{u_1, u_2, \ldots, u_n\}$. A fuzzy set defined in the universe of discourse U can be represented as:

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

where f_A denotes the membership function of the fuzzy set A, f_A: $U \to [0, 1]$, $f_A(u_i)$ denotes the degree of membership of u_i belonging to the fuzzy set A, $f_A(u_i) \in [0, 1]$, and $1 \leqslant i \leqslant n$.

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

Definition 2. Assume that $F(t)$ is caused by $F(t-1)$ only, denoted as $F(t-1) \to F(t)$, then this relationship can be expressed as $F(t) = F(t-1) \circ R(t, t-1)$, where $F(t) = F(t-1) \circ R(t, t-1)$ is called the first-order model of $F(t)$, $R(t, t-1)$ is the fuzzy relationship between $F(t-1)$ and $F(t)$, and "o" is the Max–Min composition operator.

Definition 3. Let $R(t, t-1)$ be a first-order model of $F(t)$. If for any t, $R(t, t-1) = R(t-1, t-2)$, then $F(t)$ is called a time-invariant fuzzy time-series. Otherwise, it is called a time-variant fuzzy time-series.

Definition 4. If $F(t)$ is caused by $F(t-1), F(t-2), \ldots, F(t-k)$, then the fuzzy logical relationship between them can be represented by a high-order fuzzy logical relationship. For example, a k-order fuzzy logical relationship can be expressed as follows:

$$F(t-k), \ldots, F(t-2), F(t-1) \to F(t) \tag{2}$$

Definition 5. If $F(t)$ is caused by $F(t-1), F(t-2), \ldots, F(t-k)$ and $G(t-1), G(t-2), \ldots, G(t-j)$, where $G(t)$ is another fuzzy time series. Then the fuzzy logical relationship between them is defined as a multivariable high-order fuzzy logical relationship.

Song and Chissom [1,2] established a four-step framework to manipulate the forecasting problem: (1) determine and partition the universe of discourse into intervals; (2) define fuzzy sets on the universe of discourse and fuzzify the time series; (3) derive