Information Sciences 367–368 (2016) 41–57

Contents lists available at ScienceDirect

Information Sciences

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

A novel forecasting method based on multi-order fuzzy time series and technical analysis

Furong Ye[a], Liming Zhang[b], Defu Zhang[a,*], Hamido Fujita[c], Zhiguo Gong[b]

[a] Department of Computer Science, Xiamen University, Xiamen, 361005, China
[b] Department of Computer and Information Science, University of Macau, Macau, China
[c] Faculty of Software and Information Science, Iwate Prefectural University, Iwate, Japan

ARTICLE INFO

Article history:
Received 30 November 2015
Revised 8 March 2016
Accepted 25 May 2016
Available online 30 May 2016

Keywords:
Financial forecasting
Genetic algorithm
Fuzzy time series
Technical analysis

ABSTRACT

Financial trading is one of the most common risk investment actions in the modern economic environment because financial market systems are complex non-linear dynamic systems. It is a challenge to develop the inherent rules using the traditional time series prediction technique. In this paper, we proposed a new forecasting method based on multi-order fuzzy time series, technical analysis, and a genetic algorithm. Multi-order fuzzy time series (first-order, second-order and third-order) are applied in the proposed algorithm, and to improve the performance, genetic algorithm is used to find a good domain partition. Technical analysis such as the Rate of Change (ROC), Moving Average Convergence/Divergence (MACD), and Stochastic Oscillator (KDJ) are introduced to construct multi-variable fuzzy time series, and exponential smoothing is used to eliminate noise in the time series. In addition to the root mean square error and mean square error, the directional accuracy rate (DAR) is also used in our empirical studies. We apply the proposed method to forecast five well-known stock indexes and the NTD/USD exchange rates. Experimental results demonstrate that our proposed method outperforms other existing models based on fuzzy time series.

1. Introduction

Fuzzy intervals are widely regarded as the fundamental problem for modeling fuzzy time series and are essential for model calculation and trend prediction [1]. As a result, fuzzy intervals are often taken as a key research problem in data analysis. Since Song and Chissom [1] introduced the concepts of fuzzy time series, such models have received much attention from researchers, and considerable research progress has been made afterwards. We can group the existing works into the following four categories according to the technique used to partition fuzzy intervals as follows.

As the first category, the works by Song [1,2], Chen [3], Hwang [4] and Lee [5] are regarded as the pioneer research efforts in this area. In their models, the minimum and maximum values of the sample data were rounded upward and downward, respectively, to determine the universe classification. Then, based on the size of the universe, they took an integer as the length of the interval to uniformly divide the universe.

The representative scholars of the second category include Huarng [6], Teoh [7], Jilani [8], and Yu [9]. They proposed to divide the interval based on the distribution of the sample data. The techniques used include adjusting the interval lengths

* Corresponding author. Tel.: +86 5925918207; fax: +86 5922580035.
E-mail address: dfzhang@xmu.edu.cn (D. Zhang).

http://dx.doi.org/10.1016/j.ins.2016.05.038
0020-0255/

according to the density of the samples, defining a new distance formula and dividing intervals according to the distance distribution between samples, and determining the number of intervals according to the statistical peak of the samples.

Aladag [10,11], Yolcu [12] and Egrioglu [13,14], as the representatives of the third category, proposed to find the partition method of the optimal fuzzy subset by using optimization algorithms. Aladag [10] used neural networks to forecast in high order fuzzy time series, Yolcun [12] focused on a sing-variable constraint and proposed an efficient approach to identify the intervals' length, and the method of Egrioglu is based on SARIMA. Chen and Cheng's genetic algorithm [15–17] also falls in this category. The basic idea of methods is to use the prediction error as the objective function and seek for the minimum value of the objective function according to a certain step length. The interval with the minimum objective value is taken as the final partition of the model.

The fourth category of methods includes newly proposed clustering algorithms [18–21] whose idea is similar to the fuzzy clustering method (FCM) by Li [22]. The basic idea of these methods is to use an appropriate algorithm to perform cluster analysis on the sample data, and then determine the partition of each subinterval according to the clustering results.

Many investors often use technical indicators to analyze the stock market and predict its future trend [23]. Stevenson and John [24] applied a new technique year percentage change replacing enrollments as the universe of discourse. Multivariable fuzzy time series or multi-factor fuzzy time series based on technical indicators is used to solve the problem of prediction. Lee et al. [25] developed more techniques on prediction, which considered more factors and high-order to achieve better results. To avoid complicated matrix computations, Huarng et al. [26] handled forecasting problem with a multivariate heuristic model. To forecast the TAIEX, Yu and Huarng [27,28] proposed a bivariate model by using neural networks. Chen and Chang [29] handled fuzzy rules by clustering algorithms and assigned different weights to clusters. Chen and Chen [30] investigated TAIEX forecasting with fuzzy time series by considering main factor and secondary factor. Chen et al. [31] used a particle swarm optimization algorithm to improve their research. Kim et al. [32] used technical indicators to construct multiple classifiers for predicting a stock price index. Egrioglu et al. [33] predicted Belgian traffic accident casualties from 1974 to 2004 and used a feedforward neural network to deal with fuzzy relationships in bivariate time series. Avazbeigi et al. [34] applied three variable fuzzy time series for prediction of automobile production in Iran; in addition, tabu search is presented in their work. Park et al. [35] developed bivariate fuzzy time series, which underlying price is used as the second variable, to predict TAIEX and South Korea's KOSPI 200 index.

Data processing is an important step of data mining to improve the performance of data mining algorithms. In general, data mining algorithms fail to extract nonlinear valuable patterns from noisy data; therefore, many data smoothing methods are used to address this task. By "averaging out" the noise, they can extract nonlinear relations from the time series. For example, Zhang *et al.* [36] used nonparametric kernel regression to filter the noise in the time series. This paper preprocesses the training data using the exponential smoothing method [37] to obtain smooth training data values, but do not perform any processing on the testing data.

As the global financial markets are becoming deregulated, the modeling and forecast of financial market system are becoming more complex in the risk management and derivatives rating. However, one of the key aspects of complex statistical model in financial market is accurate forecasting that could yield significant profits and could also decrease investment risks [38]. Considering the stock prediction, the most frequently used forecasting methods are nonlinear models, for example, neural network [36,39], Markov modeling [40], genetic algorithm [41], fuzzy logic [42], support vector machine [43] and hybrid models [44]. However, fuzzy time series method has been regarded as one of important novel methods in this area. Thus far, there has been various research of handling stock index forecasting using fuzzy time series [45–49].

For fuzzy time series forecasting, Chen et al. [15,16,18,19,29–31,49], Yu and Huarng [4,9,26–28,39,45] did a lot of excellent works. Recently, Wei et al. [50] forecasted the trend of TAIEX stock by combining a linear model and moving average technical index. Chen and Chen [51] introduced binning-based partition and entropy-based discretization and proposed a new fuzzy time series model based on granular computing. Chen and Chen [52] proposed a new fuzzy forecasting model, replacing the fuzzy logical relationship groups with fuzzy-trend logical relationship groups and introducing the probabilities of trends. Cai et al. [53] used ant colony optimization to obtain a good partition of the universe of discourse, and auto-regression was introduced to better use historical information. These algorithms show good ability of achieving good forecasting results.

In this paper, a new hybrid multi-order fuzzy time series model is proposed for financial forecasting, and genetic algorithm is applied to obtain good partitions of the universe of discourse. Generally, only first-order fuzzy time series is used in this area, which we think is not appropriate. Generally, stock price is influenced by historical data. The price of a certain stock on a certain day is not only related to the price of the day before but also related to the price of the near past, although they might not have the same impact strength. Obviously first-order fuzzy time series neglects the influence from the price of the near past, and this may cause the inaccuracy of forecast. To establish contact between the predicted day and the near past days, a hybrid multi-order fuzzy time series is applied. Specifically, we extract first-order, second-order and third-order fuzzy time series and average the three fuzzy values to obtain the final predicted value. We stop at the third-order because the effect of mixing higher order fuzzy time series to the results is negligible, and it affects the computation speed. Additionally, higher orders need more time series data to train a forecasting model. After many experiments, better forecasting results are achieved by using uniform weights across the orders instead of different weights. The reason in using a genetic algorithm is that the operators in the algorithm like selection, crossover, and mutation, can help the model find excellent domain partition iteratively. In particular, we introduce three indicators, the Rate of Change (ROC), Moving Average Convergence/Divergence (MACD), and Stochastic Oscillator (KDJ), to construct the multi-variables fuzzy time series,