Neurocomputing 000 (2018) 1–13

Contents lists available at ScienceDirect

Neurocomputing

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

Fuzzy time-series model based on rough set rule induction for forecasting stock price

Ching-Hsue Cheng*, Jun-He Yang

Department of Information Management, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliu, Yunlin 64002, Taiwan ROC

ARTICLE INFO

Article history:
Received 9 November 2016
Revised 15 September 2017
Accepted 3 April 2018
Available online xxx

Communicated by A. Abraham

Keywords:
Rough sets
Fuzzy time-series
Financial profits

ABSTRACT

The stock price prediction is an important issue in stock markets because it will result in significant benefits and impacts for investor. In contrast to traditional time series, fuzzy time series can solve the forecast problem with historical data of linguistic values. In order to improve forecast performance of fuzzy time-series models, this study replaced fuzzy logical relationships with rule-based algorithm to extract forecast rules from time-series observations. Therefore, this paper developed a novel fuzzy time-series model based on rough set rule induction for forecasting stock index, and this study has four contributions to improve forecast accuracy and provide investment point (in right time) to investors:

(1) Proposed a novel fuzzy time-series model to improve forecast accuracy,

(2) rough sets are employed to generate forecasting rules to replace fuzzy logical relationship rules based on the lag period,

(3) utilized adaptive expectation model to strengthen forecasting performance, and based on the meaning of adaptive parameter to observe stock fluctuation and oscillation, and

(4) proposed buy and sell rules to calculate the profit and based on three different scenarios to provide investment suggestion to investor as references.

For evaluating the proposed model, we practically collected TAIEX, Nikkei, and HSI stock price from 1998 to 2012 years as experimental dataset, and compared the listing models under three error indexes and profits criteria. The results show that the proposed method outperforms listing models in error indexes and profits.

1. Introduction

Stock investing is an exciting and challenging monetary activity, and forecasting stock trend and price plays an important role in stock market. The stock investors could have a chance to make much money in stock returns with wise decisions; however the most investors keep a pessimistic image with a heavy loss of money in stock investment. Due to stock market behavior is nonlinearity and non-stationary, and the stock price fluctuation is extremely hard to predict correctly if without have experienced or expert knowledge. Up to date, it is difficult to build a general model for forecasting stock price accurately. Nevertheless, many researchers continue to establish feasible model for approximating stock market behaviors.

Forecasting activities play an important role in our daily life; the goal of forecasting activities is to increase accuracy and profit, such as chasing predict [38] and control prediction. In financial engineering, Kuo et al. [17] have demonstrated that the general techniques used for stock market prediction are mathematical and statistical models. In time-series analysis [14], there are many time series models such as ARIMA (Autoregressive Integrated Moving Average model, [3]) and GARCH (Generalized Autoregressive Conditional Heteroskedasticity, [2]), these models have been applied to forecast stock price and trends in the financial market. However, statistical models usually deal with linear forecasting model and variables must obey statistical normal distribution for better forecasting performance. If the research data are represented by linguistic values (also named it, “linguistic intervals” such as linguistic values of age is very young, young, old) or the number of sample data is very little, the traditional forecasting methods maybe generate the bias of forecast or poor results. Therefore, many researchers have proposed different forecasting models based on fuzzy theory [37] to solve time-series problems with linguistic values.

Song and Chissom [27] first proposed a fuzzy time-series model to forecast the enrollments at University of Alabama; the fuzzy time-series model constructed the fuzzy relation R and used a

* Corresponding author.
E-mail address: chcheng@yuntech.edu.tw (C.-H. Cheng).

https://doi.org/10.1016/j.neucom.2018.04.014
0925-2312/

Max–Min composition operator to calculate forecasting values. Chen [4] proposed a fuzzy time-series model which used equal interval lengths to partition the universe of discourse and generate forecasting rules with a simplified calculation process. However, in stock forecasting, Huarng [12] extended Chen's model with additional heuristic forecasting rules to produce forecasts. In a subsequent study, Huarng proposed another model to define interval length with distribution based length and average-based length (2001). And Yu [35] proposed a weighted fuzzy time-series model with recurrent fuzzy relationships to produce forecasts. Sun et al. [30] did a prediction of Chinese stock index (CSI) future prices using fuzzy sets and multivariate fuzzy time series method. Aladağ et al. [1] proposed a partial high order Fuzzy lagged variable selection in fuzzy time series with genetic algorithms.

Because of fuzzy time-series are appropriately applied to linguistic values datasets for generating the higher accuracy. Recently, different fuzzy time-series models have been proposed to forecast nonlinear data and various applications, such as enrollment [4,27,28], temperature [6], car road accidents [13], tourism demand [18] and the stock index [5,8], etc. In the same way many researchers presented their fuzzy time-series methods to deal with the stock price forecasting.

To sum up, previous studies have four main limitations as follows: (1) most of previous studies must obey some assumptions about the variables used in the data analysis, so it is limited to be applied to all datasets; (2) most previous time series models used only one variable to forecast stock price; (3) the rules generated from ANN (Artificial neural network) are not easy to understand, and (4) VAR (Vector AutoRegression, [25]) can solve multiple variable time series problem, however there are some disadvantages of VAR: (i) the model selection procedure is complex; (ii) the model assumptions are difficult to validate; (iii) it requires a large amount of data for building model; (iv) it deals with linear model. To improve the deficiencies of previous fuzzy time-series models, and based on the advantages of LEM2 algorithm (learning from examples module, version 2, [11]), this study proposed a fuzzy time-series model based on rough set rule induction for forecasting stock index. This study has some contributions: (1) proposed a novel fuzzy time-series model to improve forecast accuracy, (2) The proposed model utilized rough set LEM2 algorithm to generate forecast rules, it's different from previous fuzzy time series using fuzzy logical relationships rule based on lag period (3) After getting the initial forecast of fuzzy time series, the adaptive expectation model is employed to strengthen forecasting performance, and observes the positive /negative value of adaptive parameter h_0 to analyze stock fluctuation and oscillation. (4) The main aim of investment is to obtain profits, this study proposed a profitable unit equation and the rules of selling and buying for investors to determine the trading time of selling and buying as references.

This rest of the paper is organized in the following. Section 2 introduces the related work such as fuzzy time-series model and rough set LEM2 algorithm. Section 3 presents the research concept and the proposed algorithm. The section 4 shows the experimental results and comparison. And the last section is conclusion.

2. The related work

In this section, fuzzy time-series model and rough set LEM2 algorithm are introduced briefly in the following.

2.1. Fuzzy time-series

In the traditional crisp set, the degree of an element belongs to a set is either one or zero. In order to deal with the uncertain and imprecise data, Zadeh [37] proposed fuzzy set theory. In thirty years ago, fuzzy set theory didn't incorporate into time-series, therefore time-series models had failed to treat the highly nonlinear and linguistic stock market behavior, until Song and Chissom [27] proposed fuzzy time-series model to forecast nonlinear data, that is, fuzzy time series model can treat the nonlinear and non-stationary problem. In recently years, there are many new fuzzy time-series models and applied in stock forecast (Sadaei et al. [31]; Rubio et al. 2017 [24]), these models usually utilized weighted fuzzy-trend or hybrid model.

Based on Song and Chissom's definitions, many fuzzy time-series models have been proposed, and these models have been successfully applied to deal with various applications. The related definitions of fuzzy time series are introduced as follows:

Definition 1. Fuzzy set

Let U be the universe of discourse, $U=\{u_1,u_2,...,u_n\}$, A fuzzy set A on U is defined as

$$A = f_A(u_1)/u_1 + f_A(u_2)/u_2 + \ldots + f_A(u_n)/u_n$$

where f_A is the membership function of fuzzy set A, f_A: $U \rightarrow [0, 1]$, u_i is an element of fuzzy set A, $f_A(u_i)$ indicates the degree of membership of u_i in A, $f_A(u_i) \in [0, 1]$ and $1 \le i \le n$.

Definition 2. Fuzzy time-series

Let $Y(t)$ ($t=\ldots, -2, -1, 0, 1, 2,...$), a subset of a real number, be the universe of discourse in which the fuzzy sets denoted as $f_i(t)$ ($i=1,2,...$) are defined, and let $F(t)$ be a collection of $f_i(t)$ ($i=1,2,...$). Then $F(t)$ is a fuzzy time-series of $Y(t)$ ($t=\ldots, -2, -1, 0, 1, 2,...$), where -2, -1 denotes one-lag period, two-lag period.

Definition 3. Fuzzy logical relationship

If there exists a fuzzy logical relationship (FLR), the relationship can be expressed as $F(t) = F(t-1) \times R(t-1, t)$, where $\times$ denotes an operation, then $F(t)$ is said to be caused by $F(t-1)$. The logical relationship between $F(t)$ and $F(t-1)$ is denoted as

$$F(t-1) \rightarrow F(t).$$

Definition 4. Fuzzy logical relationship for two consecutive fuzzy variables

Let $F(t-1)=A_i$ and $F(t)=A_j$. The relationship between two consecutive fuzzy variables, $F(t)$ and $F(t-1)$, referred to as a FLR, can be denoted by $A_i \rightarrow A_j$, where A_i is the left-hand side and A_j is the right-hand side of the FLR.

2.2. Rough set theory (RST)

RST is a nonparametric technique that has foundations in mathematical set theory and has been widely applied to decision problems [32]. RST was developed by Pawlak and Skoworn [21] and has been accepted as an effective mathematical tool for modeling vagueness and uncertainty. This approach is particularly importance in artificial intelligence research and cognitive sciences, especially in machine learning, knowledge discovery from databases, data mining, support systems, inductive reasoning and pattern recognition [16,20,22]. Recently, RST has been applied in varied fields such as medical diagnosis, drug research, process control, credit fraud detection, bankruptcy prediction, stock market rule-generation, climate change, and the development of expert systems for The NASA Space Center.

Rough set philosophy is founded on the assumption that, in the universe of discourse associated with every object, some information objects characterized by the same information are indiscernible because of insufficient information. Any set of all indiscernible objects is called an elementary set and forms a basis granule of knowledge about the universe. Any union of elementary

Please cite this article as: C.-H. Cheng, J.-H. Yang, Fuzzy time-series model based on rough set rule induction for forecasting stock price, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.04.014