Information Sciences 180 (2010) 1610–1629

Contents lists available at ScienceDirect

Information Sciences

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

A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting

Ching-Hsue Cheng [a], Tai-Liang Chen [b], Liang-Ying Wei [c,*]

[a] Department of Information Management, National Yunlin University of Science and Technology, 123, Section 3, University Road, Touliu, Yunlin 640, Taiwan
[b] Department of Information Management and Communication, Wenzao Ursuline College of Languages, 900 Mintsu, 1st Road, Kaohsiung 807, Taiwan
[c] Department of Information Management, Yuanpei University, 306 Yuanpei Street, Hsin Chu 30015, Taiwan

ARTICLE INFO

Article history:
Received 26 November 2008
Received in revised form 21 December 2009
Accepted 8 January 2010

Keywords:
Rough set theory
Genetic algorithms
Cumulative probability distribution approach
Minimize entropy principle approach
Technical indicators

ABSTRACT

In the stock market, technical analysis is a useful method for predicting stock prices. Although, professional stock analysts and fund managers usually make subjective judgments, based on objective technical indicators, it is difficult for non-professionals to apply this forecasting technique because there are too many complex technical indicators to be considered. Moreover, two drawbacks have been found in many of the past forecasting models: (1) statistical assumptions about variables are required for time series models, such as the autoregressive moving average model (ARMA) and the autoregressive conditional heteroscedasticity (ARCH), to produce forecasting models of mathematical equations, and these are not easily understood by stock investors; and (2) the rules mined from some artificial intelligence (AI) algorithms, such as neural networks (NN), are not easily realized.

In order to overcome these drawbacks, this paper proposes a hybrid forecasting model, using multi-technical indicators to predict stock price trends. Further, it includes four proposed procedures in the hybrid model to provide efficient rules for forecasting, which are evolved from the extracted rules with high support value, by using the toolset based on rough sets theory (RST): (1) select the essential technical indicators, which are highly related to the future stock price, from the popular indicators based on a correlation matrix; (2) use the cumulative probability distribution approach (CDPA) and minimize the entropy principle approach (MEPA) to partition technical indicator value and daily price fluctuation into linguistic values, based on the characteristics of the data distribution; (3) employ a RST algorithm to extract linguistic rules from the linguistic technical indicator dataset; and (4) utilize genetic algorithms (GAs) to refine the extracted rules to get better forecasting accuracy and stock return. The effectiveness of the proposed model is verified with two types of performance evaluations, accuracy and stock return, and by using a six-year period of the TAIEX (Taiwan Stock Exchange Capitalization Weighted Stock Index) as the experiment dataset. The experimental results show that the proposed model is superior to the two listed forecasting models (RST and GAs) in terms of accuracy, and the stock return evaluations have revealed that the profits produced by the proposed model are higher than the three listed models (Buy-and-Hold, RST and GAs).

* Corresponding author.
E-mail addresses: chcheng@yuntech.edu.tw (C.-H. Cheng), 97007@mail.wtuc.edu.tw (T.-L. Chen), lywei@mail.ypu.edu.tw (L.-Y. Wei).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2010.01.014

1. Introduction

In the stock market, it is very difficult to forecast stock trends because of complex factors influencing stock markets and nonlinear relationships, which are contained among different periods of stock prices. Although only a few investors profit from the stock market, millions of them still have not given up trying to make money from the market. Therefore, since the first stock market opened, numerous forecasting methods have been employed in an attempt to predict stock prices.

In the area of stock market forecasting, the technical analysis method is one of the primary analytic approaches used by investors to make investment decisions, and many researchers have been focusing on technical analysis to increase their investment returns [4,13,15]. Furthermore, the technical analysis method has the ability to forecast the future price direction by studying past market data, primarily stock price and volume. The technical analysis method assumes that stock price and volume are the two most relevant factors in determining the future direction and behavior of a particular stock or market, and that the technical indicators, coming from a mathematical formula, based on stock price and volume, can be applied to predict price fluctuations and also provide data for investors, enabling them to determine the timing for the buying or selling of stock [13].

Besides the technical analysis methods, many conventional numeric forecasting models have been proposed by financial researchers, such as Engle's [17] autoregressive conditional heteroscedasticity (ARCH) model, Bollerslev's [6] generalized ARCH (GARCH) model, Box and Jenkins' [7] autoregressive moving average (ARMA) model, and the autoregressive integrated moving average model (ARIMA).

In recent decades, many researchers have employed another approach to financial forecasting: artificial intelligence algorithms In 1990, Kinoto et al. [29] developed a prediction system for the stock market by using a neural network. Nikolopoulos and Fellrath [32] combined genetic algorithms (GAs) and a neural network to develop a hybrid expert system for investment decisions. Kim and Han [27] proposed a genetic algorithms approach in order to feature discretization and the determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. Huarng and Yu [25] applied a backpropagation neural network to establish fuzzy relationships in fuzzy time series for forecasting stock prices. Roh [42] integrated a neural network and time series model for forecasting the volatility of the stock price index.

From the literature noted above, however, three major drawbacks can be found in their forecasting methods and models: (1) stock market analysts and fund managers apply various technical indicators to forecast stock market trends, based on their personal experience, which could result in erroneous judgments of market signals; (2) for most statistical methods, there are some assumptions about the variables used in the analysis, which can not be applied to those datasets that do not follow the statistical distributions; and (3) the artificial neural network (ANN) is a black-box method, and the rules mined from it are not easily understandable.

To improve upon past forecasting models, a revised model should be able to overcome the drawbacks contained in previous models and should offer a good methodology which could be used more easily by investors. Therefore, this paper proposes a hybrid forecasting model to refine past models in stock price forecasting, and provides four novel methods in the forecasting processes: (1) select essential technical indicators by using a correlation matrix; (2) use CPDA (cumulative probability distribution approach) and MEPA (minimize entropy principle approach) to discretize condition features (technical indicators) and decision features (daily price fluctuation); (3) apply the rough set theory (RST) to produce rules from the linguistic values of technical indicators; and (4) employ genetic algorithms (GAs) to refine the extracted rules to improve forecasting accuracy and stock return.

Empirically, this paper employs two types of stock databases (stock index and individual stock price) as experimental datasets. From the model verification, it is shown that the refined processes are effective in improving forecasting accuracy, and, based on the evidence, a stock analyst or investor can employ the refined processes proposed in this paper to improve their forecasting tools or models.

The rest of this paper is organized, as follows: Section 2 introduces the related works; Section 3 demonstrates the proposed model and algorithm; Section 4 evaluates the performance of the proposed model and describes the findings; and Section 5 draws conclusions and proposes recommendations for future research.

2. Related works

This section reviews related works of technical analysis, cumulative probability distribution approach, minimize entropy principle approach, rough set theory, and genetic algorithms.

2.1. Technical analysis

Technical analysis is an attempt to predict future stock price movements by analyzing a past sequence of stock prices [39]. It relies on charts and looks for particular configurations that are supposed to have predictive value. Analysts focus on investor psychology, which represents common investors' responses to certain price formations and price movements, to analyze the fluctuations of stock market. The price at which investors are willing to buy or sell depends on personal expectation. If investors expect the security price to rise, they will buy it; if investors expect the security price to fall, they will sell it. These simple statements are the cause for a major challenge in setting security prices, because they refer to human expec-