Neurocomputing 205 (2016) 264–273

Contents lists available at ScienceDirect

Neurocomputing

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

A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock

Chung-Ho Su [a], Ching-Hsue Cheng [b,*]

[a] Department of Animation and Game Design, Shu-Te University, Kaohsiung, Taiwan
[b] Department of Information Management, National Yunlin University of Science and Technology, Yunlin, Taiwan

ARTICLE INFO

Article history:
Received 22 December 2015
Received in revised form
11 March 2016
Accepted 23 March 2016
Communicated by: Ngoc Thanh Nguyen
Available online 10 May 2016

Keywords:
Fuzzy time series
Integrated nonlinear feature selection method
Adaptive Neuro Fuzzy Inference System
Adaptive expectation model
Stock forecasting
Profit evaluation

ABSTRACT

Forecasting stock price is a hot issue for stock investors, dealers and brokers. However, it is difficult to find out the best time point to buy or sell stock, since many variables will affect the stock market, and stock dataset is time series data. Therefore many time series models have been proposed for forecasting stock price; furthermore the previous time series methods still have some problems. Hence, this paper proposes a novel ANFIS (Adaptive Neuro Fuzzy Inference System) time series model based on integrated nonlinear feature selection (INFS) method for stock forecasting. Firstly, this study proposed an integrated nonlinear feature selection method to select the important technical indicators objectively. Secondly, it used ANFIS to build time series model and test forecast performance, then utilized adaptive expectation model to strengthen the forecasting performance. In order to evaluate the performance of proposed model, the TAIEX and HSI stock market transaction data from 1998 to 2006 are collected as experimental dataset and compared with other models. The results show that the proposed method outperforms the listing models in accuracy, profit evaluation and statistical test.

1. Introduction

Forecasting stock price has become an important issue for stock investors, dealers and brokers. Nevertheless, it is difficult finding out the best time point to buy or to sell stock, because many variables must take into consideration which may affect the stock market. Stock market is one of exciting and challenging monetary activities, individual investors, stock fund managers and financial analysts who attempt to forecast stock price based on their professional knowledge or consulting professionals. Although many techniques have been developed for forecasting stock price, building an efficient stock forecasting model is still an attractive issue since even the smallest improvement in prediction accuracy can have a positive impact on investments. Therefore, stock analysts strived to discover ways which can increase the profit from the stock market.

The fluctuation of stock price, often presented nonlinear and non-stationary, and investment itself is a kind of predicted behavior for more profits, however, to build an efficient stock forecasting model is not easy. From literature review, the variables selection will affect the forecast performance, and the inappropriate method will result in larger error such that the investors will face loss. Technical analysis has become an important approach for analyzing patterns and trends of stock price [2,43]. In previous studies, selected technical indicators such as input variables depended on personal experience and opinion. Therefore, this study proposed an integrated nonlinear feature selection (INFS) method to select important variables for forecasting stock index.

In time series models, the autoregressive integrated moving average (ARIMA, [4]) is extensively utilized for constructing a forecasting model. However, many statistical methods only deal with linear forecasting model and variables must obey statistical normal distribution [12]. In order to overcome the limitation, many researchers have proposed computational intelligence techniques for financial forecasting. Kimoto et al. [27] developed a prediction system for stock market by using neural network, Roh. [38] integrated neural network into time-series model for forecasting the volatility of stock price index. Chen et al. [11] proposed a comprehensive fuzzy time-series, in which variables have linear relationships between recent periods of stock prices and fuzzy logical relationships (nonlinear relationships) mined from time-series into forecasting processes.

In the past decades, artificial neural networks (ANN) have been explored by many researchers to develop a nonlinear model for stock forecast [11,31,38] and other applications. Even though ANN showed a great deal of experimental result in many studies, the applicable

* Correspondence to: Department of Information Management, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan.
E-mail address: chcheng@yuntech.edu.tw (C.-H. Cheng).

http://dx.doi.org/10.1016/j.neucom.2016.03.068
0925-2312/

Table 1
The technical indicators are collected from the previous studies.

Indicator	Explanation
MA5	MA5 (moving average for 5 days) is the closing index of the current day [36].
MA10	MA10 (moving average for 10 days) is the closing index of the current day [36].
5BIAS	The difference between the closing value and MA5, which uses the stock price nature of returning back to average price to analyze the stock market [6].
10BIAS	The difference between the closing value and MA10, which uses the stock price nature of returning back to average price to analyze the stock market [6].
RSI	RSI compares the magnitude of recent gains to recent losses in an attempt to determine overbought and oversold conditions of an asset [6].
12PSY	PSY12 (psychological line for 12 days)$=(D_{up12}/12)*100$, D_{up12} means the number of days when price going up within 12 days [36].
10WMS%R	Williams %R is usually plotted using negative values. For the purpose of analysis and discussion, simply ignore the negative symbols. It is best to wait for the security's price to change direction before placing your trades [6].
MACD9	MACD shows the difference between a fast and slow exponential moving average (EMA) of closing prices. Fast means a short-period average, and slow means a long period one [6].
MO1	$MO1(t)=\text{price}(t)-\text{price}(t-n), n=1$ [42].
MO2	$MO2(t)=\text{price}(t)-\text{price}(t-n), n=2$ [42].
DIFN	$DIFN(t)=TAIEX(t)-NASDAQ(t)$, NASDAQ stands for NASDAQ Composite Index.
DIFF	$DIFF(t)=TAIEX(t)-TAIFEX(t)$, TAIFEX stands for Taiwan Futures Exchange.
DIFE	$DIFE(t)=TAIEX(t)-\text{DowJonesIndex}(t)$.

input variables of ANN are hard to define and select [48] and the generated rules from ANN are not easily understandable [11].

In recent years, ANFIS system has been used widely to generate nonlinear models of processes to determine input–output relationships. Therefore, ANFIS is appropriate for forecasting nonlinear financial time series and generating meaningful rules for strategizing investment tactics. Many researchers also have applied prediction techniques for financial analysis [20,39,5]. Further, many hybrid forecasting techniques have been published recently [25,30,33,35,46]. Wei [46] proposed a hybrid time series adaptive network-based fuzzy inference system (ANFIS) model to forecast stock prices.

To sum up, previous studies have four main drawbacks as follows: (1) previous researches selected important technical indicators dependent on subjective experiences and opinions; (2) most conventional methods rely upon some assumptions about the variables used in the analysis, so it is limited to be applied to all datasets; (3) most conventional time series models considered only one variable to forecast stock price; and (4) the rules generated from ANN are not easy to understand. Therefore, the purpose of this paper is to propose a novel ANFIS (Adaptive Neuro Fuzzy Inference System) time series model based on integrated nonlinear feature selection (INFS) method for stock forecasting. The contributions are: (1) proposing an integrated nonlinear feature selection method to select the important technical indicators objectively, (2) using ANFIS to build time series model and test forecast performance, (3) utilizing adaptive expectation model to strengthen the forecasting performance, and (4) verifying the proposed model with good forecasting performance.

The remaining section is organized as follows: Section 2 reviews the related literatures; Section 3 provides the framework of proposed models, and introduces the major concept and algorithm; Section 4 verifies the proposed model by two datasets and compares with other models. Finally, Section 5 represents conclusions.

2. Related works

This section briefly introduces technical indicator, time series model, subtractive clustering, adaptive network based fuzzy inference system, and adaptive expectation model.

2.1. Technical indicator

Technical indicator is forecasting the pattern of prices through the study of past stock data, primarily price and volume. For helping trade decisions in asset markets, these decisions are often generated by applying simple rules to historical price data. Technical indicator attempts to predict future stock price movements by analyzing the past sequence of stock price [37].

Analysts focus on the investor's psychology and response to certain price formation and price movements. The price depends on personal expectation which investors are willing to buy or sell. It is very important to understand that market participants anticipate future development and taking action, and their action drives the price movement. Since stock market process is highly nonlinear, many researchers have been focusing on technical indicator to improve the investment return [1,47].

A technical indicator consists in a formula that is normally applied to stock prices and volumes [19]. Technical indicator explores market information and to take into account all the necessary variables in the stock exchange information [6]. Technical indicator is one of the most popular methods in use by stock traders [2]. From the previous studies [40], the technical indicators are collected and illustrated in Table 1.

Murphy [34] summarizes the basis for technical indicator into the following three premises: (1) Market action discounts everything; (2) prices move in trends; (3) history repeats itself. Furthermore, the selection of technical and economic indicators to be used in the prediction system will depend on the following factors [15]:

(1) Availability: The data must be easily obtained.
(2) Sufficiency of the Historical Databases: There must be enough sample data for the machine learning and system testing process.
(3) Correlation of the Indicators to the Price: The data should have some relevancy to the price of the security (whether it is lagging, leading, coincidental or noise).
(4) Periodicity of the Data: The data must be available in a predictable frequency (daily, weekly, monthly, and annually).
(5) Reliability of the Data: The fast changing pace of today's global financial world and the increased in financial market volatility has resulted in difficulty to obtain reliable economic data.

2.2. Time series model

A time series is a collection of observations of well-defined data, which was obtained through repeated measurements over time. Time series analysis is a method for analyzing time series data to extract meaningful statistics and other data characteristics. Time series forecasting is to forecast future values based on previously observed values. Time series forecasting methods have