Neurocomputing 000 (2018) 1–8

Contents lists available at ScienceDirect

Neurocomputing

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

Brief papers

Predicting the direction of stock markets using optimized neural networks with Google Trends

Hongping Hu[a], Li Tang[b], Shuhua Zhang[b], Haiyan Wang[c,*]

[a] School of Science, North University of China, Taiyuan, Shanxi 030051, China
[b] Coordinated Innovation Center for Computable Modeling in Management Science, Tianjin University of Finance and Economics, Tianjin 300222, China
[c] School of Mathematical & Natural Sciences, Arizona State University, Phoenix, AZ 85069-7100, USA

ARTICLE INFO

Article history:
Received 11 July 2017
Revised 4 January 2018
Accepted 15 January 2018
Available online xxx

Communicated by Wei Chiang Hong

Keywords:
Back propagation neural network
Sine cosine algorithm
Google Trends
Stock price

ABSTRACT

The stock market is affected by many factors, such as political events, general economic conditions, and traders' expectations. Predicting the direction of stock markets movement has been one of the most widely investigated and challenging problems for investors and researchers as well. Many researchers focus on stock market analysis using advanced knowledge of mathematics, computer sciences, economics and many other disciplines. In this paper, we present an improved sine cosine algorithm (ISCA), which introduces an additional parameter into the sine cosine algorithm (SCA), to optimize the weights and basis of back propagation neural networks (BPNN). Thus, ISCA and BPNN are combined to create a new network, ISCA-BPNN, for predicting the directions of the opening stock prices for the S&P 500 and Dow Jones Industrial Average Indices, respectively. In addition, Google Trends data are taken into consideration for improving stock prediction. We analyze two types of prediction: Type I is the prediction without Google Trends and Type II is the prediction with Google Trends. The predictability of stock price direction is verified by using the hybrid ISCA-BPNN model. The experimental results indicate that ISCA–BPNN outperforms BPNN, GWO-BPNN, PSO-BPNN, WOA-BPNN and SCA-BPNN in terms of predicting the direction of the opening price for both types and significantly for Type II. The hit ratios for ISCA-BPNN with Google Trends reach 86.81% for the S&P 500 Index, and 88.98% for the Dow Jones Industrial Average Index. Our results show that Google Trends can help in predicting the direction of the stock market index.

1. Introduction

The stock market is one of the most vital components of current economy. Many people always attempt to predict stock prices including the opening prices, the highest prices, the lowest prices, the closing prices and the trading volumes of their favorite stocks. Accurately predicting the stock prices can provide investors more opportunities of gaining profit in the stock exchange [1]. But it is a difficult issue to predict stock prices. Under all kinds of influences such as economic environment, political policy, industrial development, market news, and natural factors, the stocks are dynamic and exhibit wide variation, and the prediction of the stock market thus becomes a highly challenging task because of the highly nonlinear nature and complex dimensionality [2,3].

In recent years, many techniques and various models have been researched and applied to predict the stocks. These methods include artificial neural networks (ANNs) [1,4–18], the support vector machines (SVMs) [16,19,20], the autoregressive integrated moving average (ARIMA) [6,19], the adaptive exponential smoothing [5], prospect theory [21], and multiple regression [22,23].

ANNs are nonlinear networks and have advantages of self-organizing, data-driven, self-study, self-adaptive, and associated memory similarly to human behaviors and have conducted the classification, prediction, and pattern recognition. In addition, ANNs can learn and obtain hidden functional relationships. Researchers have utilized many ANNs to predict financial time series. They include back propagation neural networks (BPNN) [1,4,10,16], elman recurrent neural network [11,24], radial basis function neural networks [12], generalized regression neural networks [13], wavelet neural networks [10,14], and so on. Unfortunately, the functional relationships between the given data and the outputs are unknown or difficult to identify.

In order to optimize the parameters of ANNs and making ANNs stable, algorithms based on population such as particle swarm optimization algorithm [15], artificial fish swarm algorithm [12], artificial bee colony algorithm (ABC) [9], fruit fly op-

* Corresponding author.
E-mail address: Haiyan.Wang@asu.edu (H. Wang).

https://doi.org/10.1016/j.neucom.2018.01.038
0925-2312/

timization algorithm [25], sine cosine algorithm (SCA) [26], and evolutionary algorithm, genetic algorithm (GA) [1,4,18,27], have been proposed.

A number of the methods above have been combined to create new models for predicting stock trends or the stock prices. For example, Qiu et al. [1] optimized the ANN model using GA to determine the optimal set of weights and biases for predicting the direction of the next days' price of the Japanese stock market index. Zahedi and Rounaghi [7] applied ANN and the principal component analysis method to predict stock price on Tehran Stock Exchange. The method can accurately predict and identify effective factors in stock price by using real values. Kumar Chandar et al. [10] used the discrete wavelet transform to decompose the financial time series data as input variables of BPNN to forecast future stock prices. Pai and Lin [19] proposed that a hybrid model of ARIMA and SVMs greatly improves the prediction performance of the single ARIMA model or the single SVMs model in forecasting stock prices. Enkea et al. [22] used a fuzzy neural network with fuzzy clustering and multiple regression analysis for stock market prediction.

The search engine, Google, provides access to aggregated information on the volume of queries for different search terms and how these volumes change over time, via the publicly available service Google Trends, which is an online search tool that allows the user to see how often specific keywords, subjects and phrases have been queried over a specific period of time. From Google Trends, people can observe similar patterns of stock spikes due to popularity of Internet. Not only does the data reflect that stock changes, but, that data can be used to anticipate certain future trends [28]. Hamid and Heiden [29] introduced an economically motivated model for using Google search frequency data to forecast volatility, based on the concept of empirical similarity (ES), and the highlight that the forecasting performance is indeed driven by the use of Google Trends data in combination with the ES framework.

In this paper, we present an improved sine cosine algorithm (ISCA) to predict the directions of the opening stock prices for the S&P 500 and Dow Jones Industrial Average (DJIA) Indices, respectively. We introduce an additional parameter into the sine cosine algorithm (SCA), to optimize the weights and basis of back propagation neural network (BPNN). In addition, Google Trends data are taken into consideration for enhancing stock prediction. We analyze two types of prediction: Type I is the prediction without Google Trends and Type II is the prediction with Google Trends. The predictability of stock price direction is verified by using the hybrid ISCA-BPNN model. The experimental results indicate that ISCA-BPNN outperforms BPNN, GWO-BPNN, PSO-BPNN, WOA-BPNN and SCA-BPNN in terms of predicting the direction of the opening stock price for both types and significantly for Type II. The hit ratios for ISCA-BPNN with Google Trends reach 86.81% for the S&P 500 Index and 88.98% for the DJIA Index, the highest among the models. Our results show that Google Trends can help in more accurately predicting the direction of the stock prices. Therefore, we improve the performance of optimizing artificial neural network (ANN) by Google Trends and an improved sine cosine algorithm based on population.

The rest of the paper is organized as followings. Section 2 introduces the related works. Section 3 proposes the improved sine cosine algorithm (ISCA) and analyzes the parameters in ISCA. We combine ISCA with the back propagation neural network (BPNN) to create ISCA-BPNN for predicting the direction of stock markets index. Section 4 presents experimental results of ISCA-BPNN based on data from Yahoo Finances and compare with other models to demonstrate the advantage of ISCA-BPNN. In addition, Google Trends data are taken into consideration for making a better stock prediction. The paper concludes with some discussions in Section 5.

2. Related works

There is a rich literature on prediction of stock markets with machine learning techniques. Here, we only choose to discuss a number of related works. One could find more in the references therein. In addition to a number of works we discuss in Section 2, Chong et al. [30] recently presented a systematic analysis of the applications of deep learning networks for stock market prediction and used deep learning networks for prediction of Korea KOSPI 38 stock returns, where four measures: normalized mean squared error (NMSE), root mean squared error (RMSE), mean absolute error (MAE), and mutual information (MI) were evaluated for prediction performance. Their results were compared to some previous results for a number of well known stock indices including the US S&P 500 [31], Korea KOSPI200 [32], US Dow Jones [33], India CNX and BSE [34], Taiwan TAIEX and US NASDAQ [35], World 22 stock market [36], Greece ASE general [37], Japan Nikkei 225 [1,18], US Apple stock [38], US SPDR S&P 500 ETF (SPY) [39], and Korea KOSPI 38 stock returns [30] Indices. Their results demonstrates that deep learning networks can be effectively used for stock predictions.

Qiu and Song [1] utilized GA to optimize BPNN for prediction of Japanese Nikkei 225 Index and used hit ratios, which are defined as the percentage of trials when the predicted direction is correct, as a criteria for predicting the direction of the stock index. They compared two basic types of input variables to predict the direction of the daily stock market index and concluded that the Type II input variables can generate a higher forecast accuracy by the use of the proposed GA-BPNN. Qiu et al. [18] also selected the input variables for BPNN for prediction of Japanese Nikkei 225 Index, and employed GA and simulated annealing (SA) to improve the prediction accuracy of BPNN and overcome the local convergence performance of BPNN. Mean Square Error (MSE) was used as a criteria for predicting the price of the stock index in [18].

Chen and Hao [40] utilized the weighted support vector machine to identify the weights and then used the feature of weighted K-nearest neighbor to effectively predict stock market indices on two well known Chinese stock market indices: Shanghai and Shenzhen stock exchange Indices. MAPE and RMSE were employed to verify the performance of the models.

Zhang et al. [41] proposed a new approach named status box method and used machine learning techniques to classify the boxes. Then they constructed a new ensemble method integrated with AdaBoost algorithm, probabilistic support vector machine (PSVM), and GA to perform the status boxes classification. They focused on Shenzhen Stock Exchange (SZSE) and National Association of Securities Dealers Automated Quotations (NASDAQ) for predicting the stock direction.

Moghaddam et al. [42] applied several feed forward ANNs with different nodes in one hidden layer or 2 hidden layers to predict NASDAQ Index based on two input datasets (four prior days and nine prior days). They constructed a different transfer function of BPNN in hidden layers and employed the values of R^2 as a criteria of models. In addition, de Oliveira et al. [43] built a neural model to predict the stock closing prices in the short term. On the other hand, The authors in Refs. [16,20,27,44] studied prediction of the stock indices in the long term and the hit ratios were employed as a criteria for measuring accuracy.

Lahmiri [45–47] combined multiresolution technique with ANN and SVR to perform training and prediction. Specifically, Lahmiri [45] used the multiresolution analysis techniques to decompose interest rates, and utilized feedforward neural network with PSO to perform prediction of interest rates. In addition, various other methods have been used for stock prediction. For example, while variational mode decomposition (VMD) has often been used as an advanced multiresolution technique for signal processing, Lahmiri [46] proposed a VMD-based generalized regression neural network