© 2019 IJSRSET | Volume 6 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology
DOI : https://doi.org/10.32628/IJSRSET

A Stock Prices Prediction Approach Via Neural Network by Several Investor Indicators

Yahui Chen[1], Zhan Wen[1*], Qi Li[2], Yuwen Pan[1], Xia Zu[1], Wenzao Li1[3]

[1]School of Communication Engineering, Chengdu University of Information Technology, Chengdu, Sichuan, China

[2]Shenwanhongyuan Securities（Sichuan Branch） Co., Chengdu, China

[3]School of Computing Science, Simon Fraser University, Burnaby BC, V5A 1S6, Canada

Corresponding Author email to * E-mail: zhanwen2005@gmail.com

ABSTRACT

The prediction of stock indicators such as prices, trends and market indices is the focus of researchers. However, stock market has the characteristics of high noise and non-linearity. Generally, linear algorithms are not good for predicting stock market indicators. Therefore, BP neural network, a model suitable for nonlinear task, is widely used in stock market forecasting. However, many BP neural network prediction models are only based on historical stock quantitative data, and do not consider the impact of investor behavior on the stock market. Therefore, based on historical stock data and quantitative data of investor behavior of ten selected Chinese stocks, this paper trains a three-layer BP neural network to predict the stock prices such as the highest price ,the opening price ,the closing price, the lowest price in a short term. And then, the model that incorporates the investor behavior indicator is compared with the model that is not added. The results show that investor behavior indicators can improve the accuracy and generalization of the stock price forecasting model effectively, especially when the model based on stock quantitative data has a poor prediction accuracy on the test set.

Keywords: BP Neural Network, Chinese Stocks, Investor Behavior And Price Prediction

I. INTRODUCTION

Since the fluctuation of the stock market can affect the income of individuals, the government and enterprises, the prediction of stock market related indicators such as stock prices, market indices and trends of individual stocks has always been a hot topic for researchers. From the 1960s to the 1970s, M. F. M Osborne and Eugene Fame et al. first studied the stock market forecasting and proposed the unpredictability of changes in the stock market[1-2]. However, in the 1990s, stock markets such as the United States showed predictability, and Cochrane's research showed that the return on investment in stocks and bonds and the foreign exchange market was predictable [3]. Later, with the development of financial theory and computer technology, researchers found that the stock market is predictable with machine learning. For instance, Rohit Choudhry and Kumkum Garg used genetic algorithm and support vector machines for stock market prediction. The price forecast accuracy has reached 61.7%[4]. However, stock market has the characteristics of high noise and non-linearity, the study of stock market forecasting using linear machine learning model can not achieve good results. Neural network is a massively parallel complex nonlinear dynamic system with nonlinear, therefore, it is very suitable for solving the problem of stock price forecasting with neural network. For example, Ticknor uses financial technical indicators and market prices as input to the ANN to predict the stock price for a certain period of time in the future, resulting in higher accuracy than linear methods[5]; Majumdar et al. established a BP neural network model to predict the direction of stock index change by 69.72% [6]; Qiu, M. et al. used ANN for the prediction of stock market returns, and got good performance in Japanese stock market[7].

But many neural networks are only based on historical stock quantitative data, such as the highest price and the opening price of stocks, etc., and do not take into account that investor behavior will also affect stock market forecasting. With the deepening of behavioral finance research, investor behavior has been proven to play a role in stock market forecasting. Therefore, based on historical stock data and quantitative data of investor behavior, this paper trains a three-layer BP neural network to predict the stock prices such as the highest price ,the opening price, the closing price, the lowest price of the ten Selected Chinese stocks. And then, the BP neural network model that incorporates the investor sentiment indicator is compared with the model that is not added.

The main contribution of this paper is to use quantitative data of investor behavior and the stocks both. This paper uses three stock indicators, which are turnover rate, psychological line and psychological average line, as the quantitative indicators of investor behavior and makes a comparison between the model with Quantitative data of investor behavior and the model that is not added in predicting Chinese stocks. To a certain extent, it proves that investor behavior has a certain impact on stock market changes, and quantitative data of investor behavior can be quantified by turnover rate, psychological line and psychological average line.

The second section (**STOCK MARKET FORECASTIONG METHODS)** will introduce the theory of stock prediction used in this paper. The third section (**EXPERIMENTAL IMPLEMENTATION**) will detail the experimental implementation. The fourth section (**CONCLUSION**) will summarize the experimental conclusions.

II. STOCK MARKET FORECASTIONG METHODS

A. Stock Market Forecast Common Indicators

The market price of a stock is affected by many factors such as composite index, price-earnings ratio, opening price, closing price, the highest price, the lowest price, turnover rate, psychological line and psychological average line et.al. In this part, we mainly introduce indicators used in this paper.

1) Opening price: refers to the price of the first transaction of the stock after the opening of the day.
2) Closing price: refers to the price of the last stock in the daily transaction.
3) The highest price: refers to the highest price among the prices traded on the day.
4) The lowest price: refers to the lowest price in the price of the day.
5) Turnover: refers to the total amount of stocks traded in currency, equal to the transaction price multiplied by the volume.
6) Turnover rate: refers to the frequency of stocks trading in the market within a certain period of time, which is one of the indicators reflecting the stock's liquidity. Studies have shown that the turnover rate can be used as an quantified variables to measure the investor attention.
7) Psychological line (Psy) and Psychological average line (Psyma): refers to an emotional response

indicator that indicates that investors have psychological changes in the stock market. It can judge the short-term trend of the stock market. Psyma is the average line of Psy.

B. Stock Forecasting Methods

In investment practice and theoretical research, there are three main methods of stock market forecasting: analytical method, linear time series method and neural network prediction method.

The analysis method is to analyze the factors that can affect the stock price to get investment decision by manual. The time series method uses the application statistics method to process a set of chronological sequence of numbers to predict the future direction. However, the above methods need manual analysis, the analyst's experience will have a different impact on the stock market analysis results to a large extent. And the stock market is a very complex nonlinear dynamic system, the linear time series model can not achieve good application results in the application of stock market forecasting.

The nonlinear neural network prediction method can select the variables or indicators in the analysis method and the time series method as the input of the model, thereby overcoming the shortcomings that the single method cannot fully reflect the market situation. And the neural network as a multi-layer prediction model that has been proved to have the function of realizing any complex nonlinear mapping is particularly suitable for use in stock market prediction related research, Among them, the single hidden layer BP neural network is the most widely used. The figure below is a block diagram of a single hidden layer BP neural network.

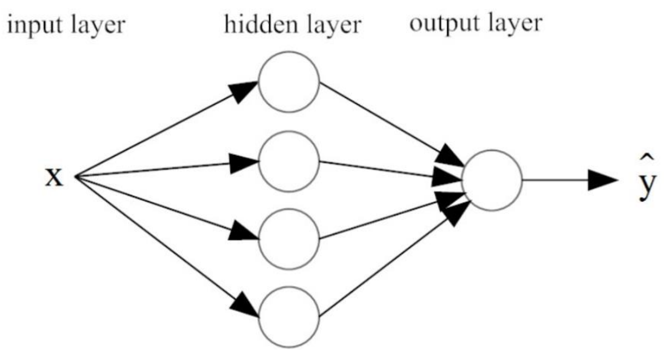


Fig. 1: a single hidden layer BP neural network

In general, the formula for a single hidden layer BP neural network is:

$$f(X) = \omega_2 g(\omega_1^T X + b_1) + b_2 \quad (1)$$

Given a set of training examples $X = \{x_1, x_2 \dots x_i\}$ where xi is the input feature. ω_1 and ω_2 represent the weights of the input layer and hidden layer. b_1 and b_2 are model parameters and represent the bias added to the hidden layer and the output layer. $g(\cdot)$ is the activation function for the hidden layer. The activation function generally has the four options: Identity, Logistic , Tanh and Relu.

Since the Relu function has a steep gradient, the speed of the gradient can be accelerated when the weight is adjusted in the gradient descent algorithm, so the Relu function is widely used in the prediction problem. The formula is:

$$f(x) = max(0, x) \quad (2)$$

x is the input variables, compare it to 0 and choose the maximum. Because of the higher speed in the gradient descent duration, Relu is used in this paper.

Starting from initial random weights, the model minimizes the loss function by repeatedly updating these weights. After computing the loss, a backward pass propagates it from the output layer to the previous layers, providing each weight parameter with an update value meant to decrease the loss.

Because the SGD optimizer is prone to the problem of slow gradient change and local optimality during the gradient descent, this paper used the Adam optimization algorithm which is similar to SGD that can automatically adjust the learning rate to solve the above problem[8-9]. The algorithm stops when it reaches a preset maximum number of iterations; or when the improvement in loss is below a certain, small number.

C. **Behavioral Finance**

Since the behavior of investors with subjective initiative is complex, the simple use of stock quantitative data cannot fully predict the changes in the stock market. Therefore, researchers have introduced the concept of behavioral finance in stock forecasting, that is, investors' decision-making behavior will have a certain impact on stock market.

Investor behavior is divided into two aspects: investor attention and investor sentiment. Investor attention can be measured by using proxy variables that indirectly reflect investor concerns about the stock market or direct proxy variable like web search data. For example, Da et al. used Google Trends search indicators to conduct research, and succeeded in explaining the short-term rise in stock prices and the stock price reversal effect[10]; as Dimpfl et al.'s research, there is a strong common performance between online search data such as the volatility of search index and stock market index in Google or Baidu trend[11].

The research on extracting investor sentiment data from stock market related data officially appeared in 1991. Researchers such as Lee formally proposed the concept of investor sentiment, and used the volatility of closed-end fund discount as a proxy variable of investor sentiment[12]; Barker et al.'s research indicates that agents such as turnover rate, IPO first-day yield, and closed-end fund discount rate are used to represent investment emotion, the effect of predicting stock returns is good [13];

Therefore, from the perspective of theoretical modeling and empirical research, investor behavior has a strong correlation with stock market volatility. The influence factors of investor sentiment and attention are of great significance in stock market forecasting.

III. EXPERIMENTAL IMPLEMENTATION

A. **Data Preprocessing**

This paper selects a total of 14421 Chinese stocks with stable state and no abnormal fluctuations between 2013 and 2018. The names, codes and quantities of the ten stocks are shown in the table below.

TABLE 1

TEN SELECTED CHINESE STOCKS

Stock name	Stock code	Select quantity
Baiyun Airport	sh600004	1457
Baotou Steel	sh600010	1406
Dongfeng Motor	sh600006	1407
Huaneng International	sh600011	1457
HSBC Bank	sh600015	1457
Minsheng Bank	sh600016	1457
Pufa Bank	sh600000	1424
Rizhao Port	sh600017	1457
Capital	sh600008	1442
China International Trade	sh600007	1457

From the common stock market indicators, this paper selects 8 indicators such as opening price, closing price, highest price, lowest price, turnover, turnover rate, psychological line and psychological average as the input characteristics of BP neural network. The opening price, closing price, highest price, lowest price and turnover amount are financial quantitative data, which are input characteristics of BP neural network based on financial quantitative data. The turnover rate, psychological line and psychological average are used as investor sentiment and investor attention agent as input characteristics of BP neural network which is a quantitative indicator of investor behavior.

The experiment uses a one-step forecasting method that uses a five-day cycle to predict the stock price for the next day. Therefore, the stock vector based on financial quantitative data predicts that an input vector of BP neural network has 25 input features, and an input vector of neural network added to investor sentiment has 40 input features. The above-mentioned neural network model predicts the stock price of the next day, which is the opening price, closing price, highest price and lowest price of individual stocks.

The data collected in this experiment was downloaded from the forecaster website.

B. BP Neural Network Based on Financial Quantitative Data

In the BP neural network based on financial quantitative data realized in this section, the input feature is the financial quantitative data of the selected stocks, namely, the opening price, the closing price, the highest price, the lowest price, and the turnover amount. The input data is in a single-step forecast, five days is a cycle, and the opening, closing, highest, and lowest prices of the selected stocks are selected one day after the forecast. The model activation function selects Relu, and the optimizer selects the Adma optimizer.

According to the kolmogrov theorem, the number of neurons in the neural network should be twice the number of input features plus one, and the theoretical optimal number of nodes in the neural network is 51, so the nerual network of 49-52 nodes are selected during the experiment. The rest of the parameters are set as shown in the table below.

TABLE 2

PARAMETERS OF THE MODEL

Name of parameters	Value of parameters
Alpha	0.0001
Learning_rate	Adaptive
Learning_rate_init	0.001
Max_iter	200
Beta_1	0.9,
Beta_2	0.999

Alpha is a regular penalty factor that is added to avoid overfitting. The learning_rate is set to adaptive, the initial value is set to 0.001 by experience, and the maximum number of iterations is 200. Beta_1 and beta_2 are parameters for ADAM to adjust each layer of the model parameters, and are set to 0.9 and 0.999 according to experience.

In this experiment, a 5-fold cross-validation was performed to test whether the test results had over-fitting problems. The model score is measured by the R^2 standard. The closer the score is to 1, the higher the model accuracy. During the experiment, it is found that when the number of nodes is 51, the model get the best scores of predicting the opening price, closing price, highest price and lowest price. The following is the average score of the training set and the test set of ten stocks when the number of nodes is 51.

TABLE 3. THE AVERAGE SCORE OF THE TEN STOCKS

Code	Open _train	Open _test	Close _train	Close _test	High _train	High _test	Low _train	Low _test
sh600004	0.9891	0.7827	0.9849	0.7152	0.9873	0.7612	0.9861	0.7161
sh600010	0.9876	0.8685	0.9830	0.8320	0.9867	0.8510	0.9847	0.8625
sh600006	0.9859	0.3278	0.9837	0.3010	0.9882	0.3773	0.9839	0.1546
sh600011	0.9893	0.9147	0.9824	0.8668	0.9870	0.8907	0.9852	0.8702
sh600015	0.9859	0.9458	0.9822	0.9336	0.9857	0.9445	0.9836	0.9336
sh600016	0.9732	0.8925	0.9805	0.8989	0.9848	0.9301	0.9698	0.8629
sh600000	0.9890	0.8699	0.9871	0.8582	0.9880	0.8650	0.9870	0.8434
sh600017	0.9889	0.7521	0.9865	0.6944	0.9888	0.7038	0.9860	0.6814
sh600008	0.9876	0.5216	0.9853	0.5696	0.9875	0.5288	0.9860	0.5480
sh600007	0.9872	0.8360	0.9831	0.7950	0.9860	0.8366	0.9862	0.8133

C. BP Neural Network with Quantitative Data of Investor Behavior

In this section, based on the BP neural network based on financial quantitative data, the quantitative data of investor behavior is added. The input features are the opening price, closing price, highest price, lowest price, turnover, turnover rate, psychological line and psychological average line of selected stocks. Among them, the turnover rate as an investor's attention to the measurement index, the psychological line and the psychological average as a quantitative indicator of investor sentiment. The forecasting method, activation function and the optimizer are consistent with the previous section. According to the kolmogrov theorem, the theoretical optimal number of nodes of the neural network is 81, so during the experiment, the neural network with 79-82 nodes was selected and the rest of the parameters were the same as the previous one.

In this experiment, a 5-fold cross-validation was performed to test whether the test results had over-fitting problems. The model score is measured by the R^2 standard. The closer the score is to 1, the higher the model accuracy.

From the experimental results, it is concluded that when the number of nodes is 81, the model scores the best when predicting the opening price, closing price, highest price and lowest price. Therefore, the number of nodes is 81, and a single hidden layer BP neural network with investment behavior is established. The following is the average score of the training set and the test set when the number of nodes is 81, the predicted opening price, closing price, highest price, and lowest price of ten stocks.

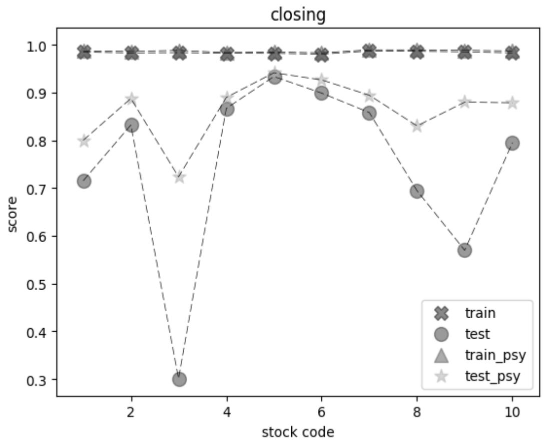
TABLE 4. THE AVERAGE SCORE OF THE TEN STOCKS

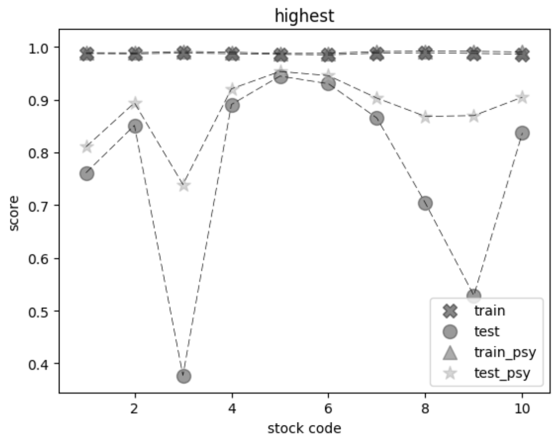
Code	Open _train	Open _test	Close _train	Close _test	High _train	High _test	Low _train	Low _test
sh600004	0.9908	0.8271	0.9873	0.7999	0.9887	0.8103	0.9881	0.8011
sh600010	0.9890	0.8922	0.9862	0.8877	0.9886	0.8938	0.9863	0.8886
sh600006	0.9886	0.6745	0.9884	0.7234	0.9911	0.7390	0.9881	0.6431
sh600011	0.9903	0.9326	0.9838	0.8898	0.9897	0.9204	0.9863	0.8990
sh600015	0.9875	0.9511	0.9851	0.9415	0.9874	0.9533	0.9860	0.9376
sh600016	0.9899	0.9548	0.9832	0.9265	0.9877	0.9455	0.9857	0.9282
sh600000	0.9915	0.9103	0.9895	0.8944	0.9910	0.9028	0.9902	0.8905
sh600017	0.9919	0.8643	0.9893	0.8302	0.9918	0.8681	0.9896	0.8296
sh600008	0.9911	0.8902	0.9887	0.8800	0.9915	0.8696	0.9887	0.8957
sh600007	0.9910	0.9088	0.9871	0.8786	0.9901	0.9045	0.9885	0.8797

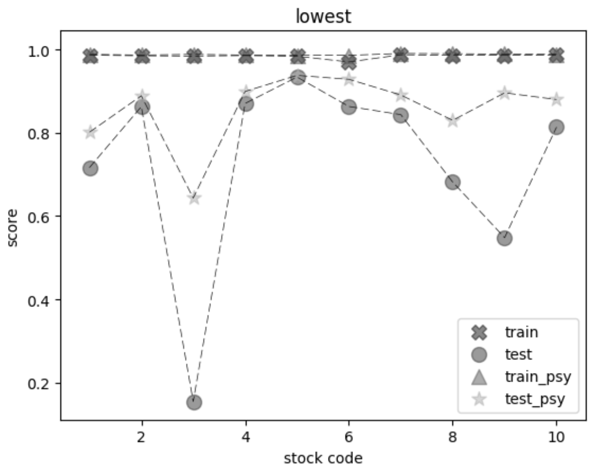
D. Model Comparison

In this section, the BP neural network model based on financial quantitative data and the BP neural network model with quantitative data of investor behavior are compared.

Below, under the optimal number of nodes, two neural network model's scores of ten stocks.







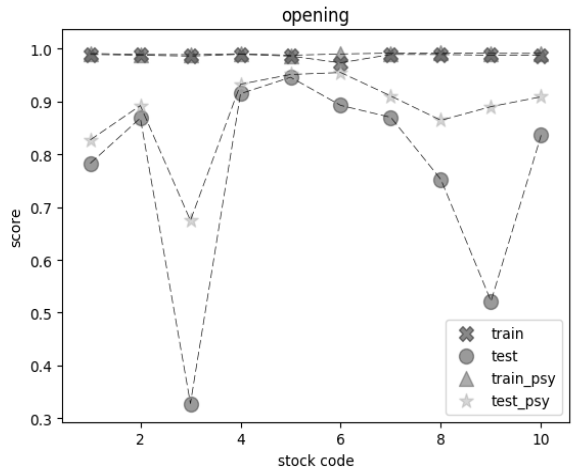


Fig.2 : The prediction results of ten stocks' the lowest prices, opening prices, closing prices and the highest prices

The train line and the test line in the figure represent the scores of the BP neural network based on financial quantitative data on the test set and the training set when the number of nodes is 51. The train_psy line and test_psy line in the figure represent the scores of the BP neural network added to the investor behavior quantitative data on the test set and the training set when the number of nodes is 81. It can be seen that the prediction effect is improved both in the training set and the test set. And when the BP neural network model based on financial quantitative data has a very poor prediction effect on the test set, the prediction effect of the BP neural network with investor behavior on the test set will be greatly improved.

III.CONCLUSION

The BP neural network that joins the investor behavior compared with the BP neural network using only financial quantitative data has significantly improved the Predictive effect of opening, closing, the highest and the lowest price of stocks. Especially when the BP neural network based on financial quantitative data has a poor prediction effect on the test set, adding the investor behavior quantitative index can greatly improve the prediction effect of the model on the test set. This paper proves that the three indicators of turnover rate, psychological line and psychological average can improve the accuracy and generalization of the model, and prove that the quantitative indicators of investor behavior can improve the model prediction effect.

IV. ACKNOWLEDGEMENTS

This research is supported by Science and Technology Department of Sichuan Province, Fund of Science and Technology Planning (No. 2018JY0290), Meteorological Information and Signal Processing Key Laboratory of Sichuan Higher Education Institutes (No. QXXCSYS201606), and Humanity and Social Science Research Project of Ministry of Education, Youth Foundation (17YJC190035).

V. REFERENCES

[1]. Brownian Motion in the Stock Market Operations Research, Vol. 7 (1959), pp. 145-173 by M. F. M. Osborne

[2]. EUGENE F. FAMA. The Behavior of Stock-Market PricesJ]. Journal of Business 38, 34-105, 1965

[3]. John H. Cochrane. New facts in financeJ]. Economic Perspectives Federal Reserve Bank of Chicago ,1999,(23):36-58.

[4]. Choudhry, Rohit & Garg, Kumkum. (2008). A Hybrid Machine Learning System for Stock Market Forecasting. World Academy of Science, Engineering and Technology. 39.

[5]. Ticknor J L. A Bayesian regularized artificial neural network for stock market forecasting J]. Expert Systems with Applications, 2013, 40(14):5501-5506.

[6]. Majumdar S, Hussein D. Forecasting of Indian Stock Market Index Using Artificial Neural OL]. Network, 2010.

[7]. Qiu, M., Song, Y., & Akagi, F. (2016). Application of artificial neural network for the prediction of stock market returns The case of the Japanese stock market. Chaos, Solitons & Fractals, 85, 1–7.pdf

[8]. Diederik P. KingmaJimmy Lei Ba. Adam: a Method for Stochastic Optimization International Conference On Learning Representations ,2015.

[9]. J. Konečný, J. Liu, P. Richtárik and M. Takáč, "Mini-Batch Semi-Stochastic Gradient Descent in the Proximal Setting," in IEEE Journal of Selected Topics in Signal Processing, vol. 10, no. 2, pp. 242-255, March 2016.

[10]. Da Z, Engleberg J, Gao P. In Search of attentionJ]. The Journal of Finance, 2011, 66(5):1461-1499.

[11]. Thomas D. and S. Jank. Can Internet Search Queries Help to Predict Stock Market Volatility?J]. Working Paper Series. 2011.

[12]. Lee C, Shellfire A, Taller R. Investor Sentiment and the Closed-end Foud PuzzleJ]. Journal of Finance, 1991, 46（1） : 75-109.

[13]. Barker M, Warbler J. Investor Sentiment and Cross-section of Stock Returns J]. Journal of Finance, 2006, 61（4） : 1645-1680.

Cite this article as :

Yahui Chen, Zhan Wen, Qi Li, Yuwen Pan, Xia Zu, Wenzao Li1, "A Stock Prices Prediction Approach Via Neural Network by Several Investor Indicators", International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET), Online ISSN : 2394-4099, Print ISSN : 2395-1990, Volume 6 Issue 2, pp. 477-484, March-April 2019. Available at doi : https://doi.org/10.32628/IJSRSET1962126
Journal URL : http://ijsrset.com/IJSRSET1962126