Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 17, No. 1, 2024

Stock Price Prediction Based on BP Neural Network and ARIMA Model

Ruiqi Zhao

Nanjing University of Posts and Telecommunications, Nanjing, China

Abstract: With the progress and development of society, stock investment has become an important part of people's daily lives. High risk and high return are the characteristics of stock investment. Individual investors and institutional investors always pay attention to stock market trends, analyze financial data, and try to predict the development trend of stocks. Studying the future trend of stock prices of listed companies and predicting the stock price of a certain company not only has extremely attractive application value, but also has significant theoretical significance, attracting the attention of investors and academia. The stock market is unpredictable and an extremely complex system with too many influencing factors. It will be very difficult to thoroughly understand the mechanism of stock market changes theoretically. This study investigates the application of time series prediction technology in stock price prediction, proposes stock price prediction models based on ARIMA and BP neural network, and verifies the feasibility and effectiveness of the model prediction through prediction.

Keywords: ARIMA, BP neural network, Stock price prediction.

1. Introduction

1.1. Purpose and Significance of Stock Price Prediction

The stock market holds an important position in the field of financial investment. With the development of social economy, people are paying more and more attention to financial management and investment. The stock market has gradually become an equally important investment channel as banks and insurance, and is one of the main ways for many families and individuals to manage their finances. In recent years, with the continuous rise of the stock market, the number of investors buying into the Shanghai and Shenzhen stock markets has been increasing. According to statistics from a Chinese securities company, at the beginning of each year, nearly 9000 new users are added in a single day. At present, the total number of users in the Shanghai and Shenzhen stock markets far exceeds 75 million. When the market is at a low point, some relatively rational institutional investors with large amounts of funds are often able to seize the opportunity to enter the market early and build positions calmly. However, by the time the stock market is at its craziest, they are already busy picking the fruits. In this regard, retail investors should learn to analyze and invest in the stock market with a longer-term perspective. However, due to factors such as inaccurate information, differences in investor perceptions, the complexity of various analytical techniques, and the randomness of stock price changes, actual investments often fail to achieve, or even the opposite, expected results, resulting in the loss of investors' capital, known as stock market risk. The purpose of price prediction is to try to find a method that can achieve maximum returns with minimal risk. Since the emergence of stocks, there have been countless people trying to predict prices. It can be said that every investor is looking for a method to predict stock prices. Many scholars at home and abroad are dedicated to studying the changing trends of the stock market and establishing corresponding prediction models, providing prediction methods, striving to avoid large stock market fluctuations, reduce investment risks, and maintain economic prosperity and stability. However, although there are countless people studying and predicting stock prices, almost no one can find a perfect and foolproof method to predict stock prices. As a foreign economist said, stock market changes are like walking a drunken man, difficult to predict. Artificial neural networks are a rapidly developing interdisciplinary field widely used in signal processing, adaptive control, and financial prediction.

1.2. Current Methods for Predicting Stock Prices

1.2.1. Decision Tree Model

Decision tree is also known as decision tree. In the decision tree method, the decision tree is first constructed from the instance set, which is a guided learning approach. This method first forms a decision tree based on the training set data. A decision tree represents the tree structure of a decision set. The final result is a tree, with leaf nodes being class names and middle nodes being attributes with branches corresponding to a possible value of the attribute. The core of decision tree technology in discovering data patterns and rules is the inductive algorithm. Inductive reasoning derives a regular conclusion from the characteristics, features, or properties represented by several facts through comparison, summarization, and generalization. Inductive reasoning attempts to obtain a complete and correct description from specific observations of a part or the whole of an object, that is, to draw universal conclusions from specific facts.

1.2.2. Neural Network Model

Neural networks are data models that simulate the structure of the human brain; therefore, neural networks are systems with self-learning capabilities. Just like the brain, neural networks learn from a set of input data and adjust model parameters based on this new cognition to discover patterns in the data. Therefore, the working process of neural networks can be divided into two stages.

Learning stage: Training the network, mainly adjusting the connection weights and connections between neurons in the network. The information processing capability of neural

networks is mainly determined by the connection method and connection weights. Because the information processing capability of neural networks is mainly acquired through the learning stage and affects the entire working stage, the connection method and connection weights between neurons in neural networks are also known as long-term memory. Although the time required for different learning modes and algorithms in neural networks varies, generally speaking, the training time of neural networks is longer and much longer than the processing time of individual data.

Working stage: The trained network can be used for practical work. At this time, the connection weights and connection parties of the network are fixed and unchanged. The working process is manifested as the mapping and change of input data in the state space. The final stable state of the neural network is the working output. Compared to the time cost of the learning phase, the speed of the work phase is relatively fast. Because the state of neurons often changes during the working stage, the state of neurons in a single stage is called short-term memory.

1.3. Application of Neural Network Theory in Stock Price Prediction

Due to the high degree of uncertainty and unpredictability in the stock market, as well as the strong non-linear prediction ability of neural network theory, many people are trying to use neural network theory to predict stock prices in order to grasp the laws of the stock market, and have achieved satisfactory results in different directions. Below are some methods of using neural network theory for prediction.

The significance and practical application value of this research project lie in:

(1) Explore new stock investment risk analysis and evaluation techniques, enrich and improve the system of stock investment risk analysis and evaluation methods.

(2) Introducing data mining techniques into stock investment risk analysis and evaluation is beneficial for promoting the development of methods for stock investment risk analysis and evaluation.

(3) Providing a quantitative analysis technique for stock investment risk based on multiple factors and applying multiple evaluation and prediction index indicators is beneficial for improving the accuracy of evaluations.

(4) Provide new ideas and practical methods for the investment activities and decisions of individual and institutional investors in the stock market.

2. Stock Price Prediction Based on ARIMA Model

2.1. Main Technical Introduction

2.1.1. Mean Square Error of MSE

MSE is used to measure the deviation between predicted and true values, and its formula is as follows:

$$\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2 \quad (1)$$

In equation (1), is the predicted value, is the true value, and m is the data size.

2.1.2. ADF Unit Root Inspection

Unit root test refers to checking whether there are unit roots in a sequence, as the existence of unit roots is a non-stationary time series. Unit root refers to the unit root process, which can prove that the presence of unit root processes in a sequence is non-stationary and can lead to spurious regression in regression analysis. Stability is a necessary condition for the autoregressive model ARMA, so for time series, the first step is to ensure that the n-order difference sequence applied to autoregression is stationary.

2.1.3. ACF, PACF autocorrelation coefficient and partial autocorrelation coefficient

In time series analysis, autocorrelation function (ACF) and partial autocorrelation function (PACF) are commonly used to determine the coefficients and order of ARMA (p, q) models. The autocorrelation function (ACF) describes the linear correlation between time series observations and their past observations. Partial autocorrelation function (PACF) describes the linear correlation between time series observations and their past observations given intermediate observations. The calculation formula is as follows:

$$\mathrm{ACF(k)} = \rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)} \quad (2)$$

2.1.4. AIC Criteria

AIC: Akaike Information Criterion (AIC) is a commonly used method for evaluating the quality of ARIMA models. Its calculation formula is as follows:

$$\mathrm{AIC} = 2\mathrm{k} - 2\ln(\mathrm{L}) \quad (3)$$

Among them, k is the number of model parameters, n is the number of samples, and L is the likelihood function.

2.1.5. ARIMA Model

(1) Autoregressive model AR

The autoregressive model first needs to determine an order p, which represents the number of historical values used to predict the current value. The formula for the p-order autoregressive model is defined as:

$$y_t = \mu + \sum_{i=1}^{p}\gamma_i y_{t-i} + \epsilon_t \quad (4)$$

In the above equation, yt is the current value, u is a constant term, p is the order, γ i is the autocorrelation coefficient, and Et is the error.

(2) Moving Average Model MA

The moving average model focuses on the accumulation of error terms in the autoregressive model. The formula for the q-order autoregressive process is defined as follows:

$$y_t = \mu + \epsilon_t + \sum_{i=1}^{q}\theta_i \epsilon_{t-i} \quad (5)$$

The moving average method can effectively eliminate random fluctuations in predictions.

(3) Autoregressive Moving Average Model ARMA

By combining the autoregressive model AR with the moving average model MA, we obtain the autoregressive moving average model ARMA (p, q), which is calculated using the following formula:

$$y_t = \mu + \sum_{i=1}^{p}\gamma_i y_{i-1} + \epsilon_t + \sum_{i=1}^{q}\theta_i \epsilon_{t-i} \quad (6)$$

2.2. Establishment, Analysis, and Validation of Time Series Data Models

The ARIMA model is used in this article, and the selected data includes the closing, opening, and fluctuations of a listed company from March 26, 2020 to March 26, 2021.

Using ARIMA (p, d, q) model to predict stock prices can help us better understand financial related knowledge.

The stock prices of listed companies can be predicted using the basic ARIMA (Automatic Regression Integral Moving Average) model.

2.3. Implementation Steps

1) Load data and extract its stable data. If the data is not stable, differential operations can be performed until the data becomes stable.

2) Conduct hypothesis testing, and if successful, continue with (h=logic 1)

3) Generate autocorrelation and partial autocorrelation that help find AR and MA ranges

4) Construct algorithms to find the best ARIMA model

5) Use the selected ARIMA model to generate prediction curves and compare them with the original data

6) Predict the future stock price and compare it with the actual value.

Organize the raw data and create a trend chart as shown below, which provides a preliminary understanding of the stock's upward trend from March 2020 to March 2021.

Figure 1. Original Data

Stability analysis was conducted on the data. As shown in Figure 2, the sequence fluctuated randomly around 0, with stable fluctuations and no obvious trend changes. The data is a stationary time series.

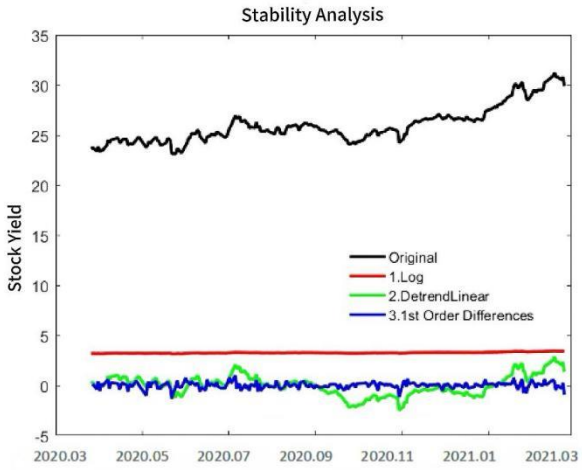

Figure 2. Stability Analysis

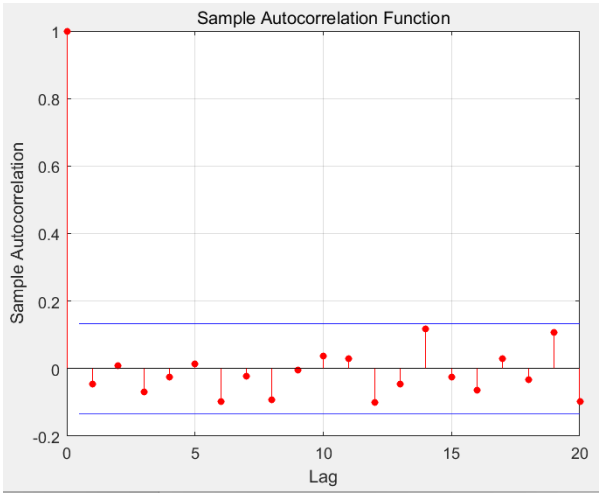

Figure 3. Autocorrelation Coefficient

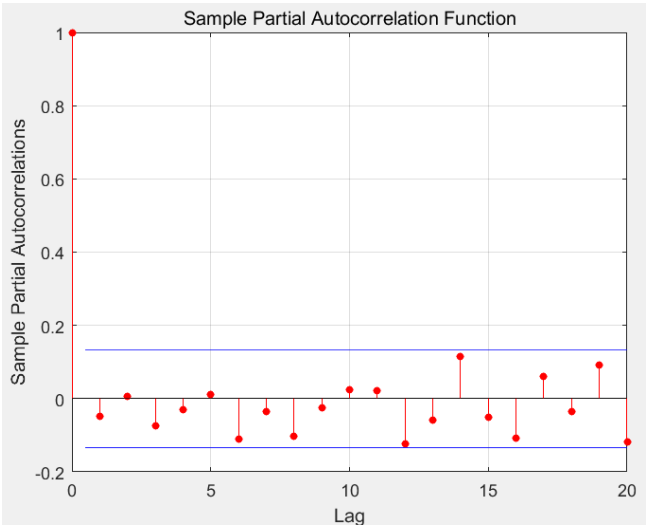

Figure 4. Partial autocorrelation coefficient

Draw autocorrelation and partial autocorrelation graphs. From Figures 4 and 5, it can be seen that the 1st and 2nd order partial autocorrelation coefficients of the PAC sequence exceed ± 2 times the estimated standard deviation. After the 2nd order, the partial autocorrelation coefficient is within ± 2 times the estimated standard deviation and rapidly decreases to 0, indicating that the partial autocorrelation function is truncated after the 2nd order; Similarly, if the correlation coefficient of the PAC sequence exceeds 5% and falls outside ± 2 times the estimated standard deviation, that is, the autocorrelation function is rounded off, and combined with the table, it can be preliminarily determined that p=1 or 2 and q=0. Using the AIC criterion to traverse the minimum AIC value yields p=7 and q=5.

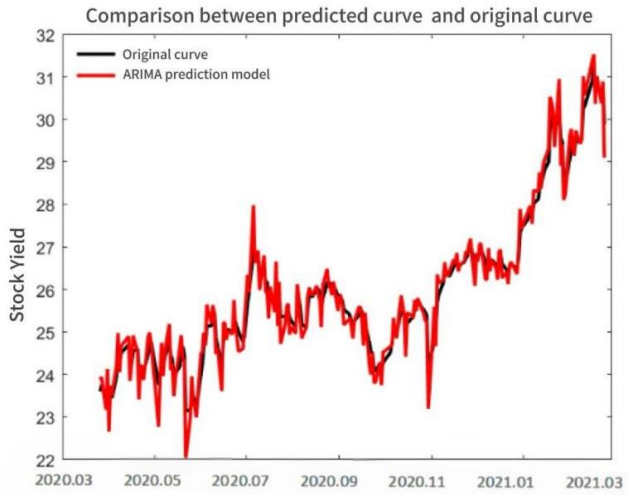

Figure 5. Predicted and Original Values

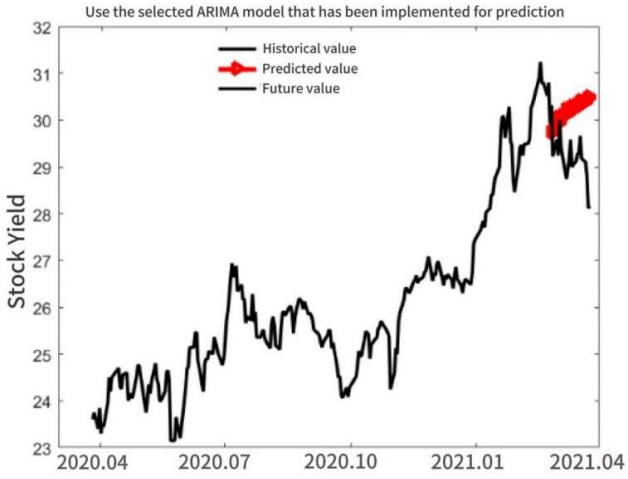

Figure 6. Prediction

Finally, the predicted data output by Matlab is shown below:

```
ans =

   29.7610
   29.7474
   29.9148
   30.0515
   29.9535
   30.0314
   30.0936
   30.2367
   30.2234
   30.1562
   30.2059
   30.2901
   30.3134
   30.2699
   30.3073
   30.3836
   30.4015
   30.3951
   30.4313
   30.4853
   30.4995
```

Figure 7. Predicted Data

3. Stock Data Prediction Model Based on BP Network

3.1. Algorithm Introduction

BP (Back Propagation) neural network was proposed in 1986 by a team of scientists led by Rumelhart and McColland. It is a multi-layer feedforward network trained using error backpropagation algorithm. BP network can learn and store a large number of input-output pattern mapping relationships without revealing the mathematical equations that describe these mapping relationships in advance. Its learning rule is to use the steepest descent method to continuously adjust the weights and thresholds of the network through backpropagation, in order to minimize the sum of squared errors of the network. It is currently one of the most widely used neural network models. Therefore, understanding the structure of BP networks and weight adjustment algorithms plays an important role in learning other neural networks. BP neural network is mainly applied in the following aspects.

Function approximation: Train a network to approximate a function using input vectors and corresponding output vectors.

Pattern recognition: Connect it to the input vector using a specific output vector.

Classification: Classify the appropriate way defined by the input vector.

Data compression: Reduce the dimensionality of output vectors for transmission or storage.

3.2. Specific Steps of BP Neural Network

3.2.1. Determination of network layers and ANN input nodes

This article adopts a three-layer network model. The input layer serves as a buffer memory, accepting external input data, so the number of nodes depends on the dimensionality of the input vector. The stock price trend follows a wave like pattern and exhibits periodicity. In technical analysis, the appropriate choice of analysis cycle has a direct impact on the prediction results. Due to time, task volume, and ease of selection, we have chosen 1 day, which means date=1.

3.2.2. Data preprocessing

Data preprocessing is the process of converting data obtained from the stock market into input data that can be recognized by ANN. Starting from day n, let ANN predict the rise and fall of stock prices. The raw data sequence of the most recently traded stocks $\{x'(t)\}$, as well as the processed data sequence $\{x(t)\}$ input to ANN, should be processed as follows:

$$x' = date - 1i = 0!x'(n-1)date, x(t) = x'(t) - x'x' + 0.5$$

3.2.3. Number of nodes in the hidden layer

The number of hidden layer nodes is directly related to the required number of input-output units for solving the problem. For BP networks, the following formula can be used for design:

$$n = n_i + n_0 + a$$

In the formula, n represents the number of hidden layer nodes; n_i is the number of input nodes; n_0 is the number of output nodes; a is a constant between 1 and 10.

3.3. Output Node Definition

The output node number of the artificial neural network is 1, and the threshold a=0 for the trend of stock price changes is selected as the average value of the rise and fall over a certain period of time; Prediction step size k=1. The output node Y (t) can be defined as follows:

When Y (t)=0, it is predicted that the daily increase or decrease of the stock price starting from t+1 will not exceed a, indicating that it is expected to be in a downward trend;

When Y (t)=1, it is expected that the stock price will be on an upward trend in the single day starting from t+1, with a cumulative increase exceeding a.

3.4. Determination of Other Network Parameters

The number of hidden layer nodes is set to 6, and the node action function is an S function, that is, f (s)=11+e-s; The learning coefficient is 0.7. When learning, it is required that the output error d be ≤ 0.1. When predicting, if the output of the output layer is greater than 0.5, it is judged as 1; Less than or equal to 0.5 is judged as 0.

The data selected for this project is the closing, opening, and price fluctuations of Swire Group A from January 3, 2017 to December 31, 2019. Four sets of data, including opening, highest, lowest, and closing, are selected to predict the closing data of the next day using the five sets of data from that day, thus establishing an equivalent stock data prediction model. Adopting a three-layer BP network structure consisting of an input layer, a hidden layer, and an output layer, the input layer contains four neurons, the hidden layer contains three neurons, and the output layer is one neuron. Among them, the activation function of the hidden layer neurons adopts an asymmetric Sigmoid function, and the function expression is:. The activation function of the output layer neurons adopts a linear function. The 738 sets of data are divided into three equal parts, of which two parts are used as training samples to train and learn the network; Another sample is used as a test to evaluate the generalization ability of the trained network. The BP algorithm is used to modify the weights of the hidden layer and output layer, in order to minimize the difference between the calculated output and the actual sample output, and ultimately achieve more accurate prediction.

Results and Analysis

Figure 8 shows the display content of the MATLAB command window, Figure 9 shows the error curve of the training sample set, Figure 10 shows the comparison between the calculated and actual outputs of the training sample set, and Figure 11 shows the comparison between the predicted and actual data of the test sample set.

```
W1 =

   15.2341    8.2518  117.3235   60.5994
   17.2322   16.8558   43.1180   58.5363
   14.9673    7.6432  118.5993   60.8658


W2 =

   51.3105  -27.2409   52.6996

Test sample set error
    2.2473
```

Figure 8. MATLAB Command Window Display Content

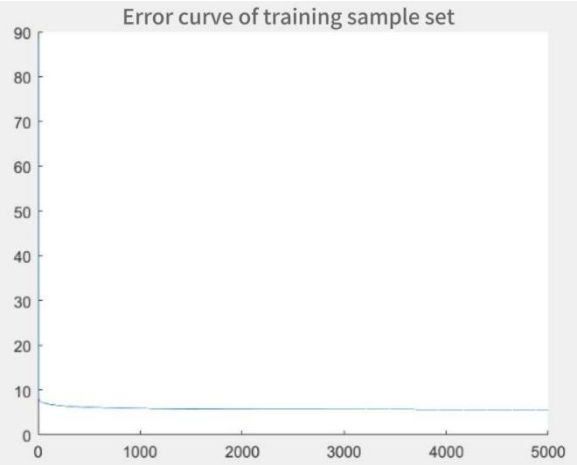

Figure 9. Error Curve of Training Sample Set

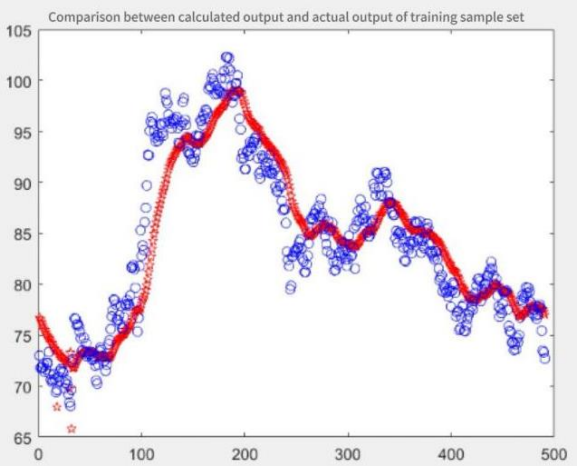

Figure 10. Comparison between Calculated Output and Actual Output of Training Sample Set

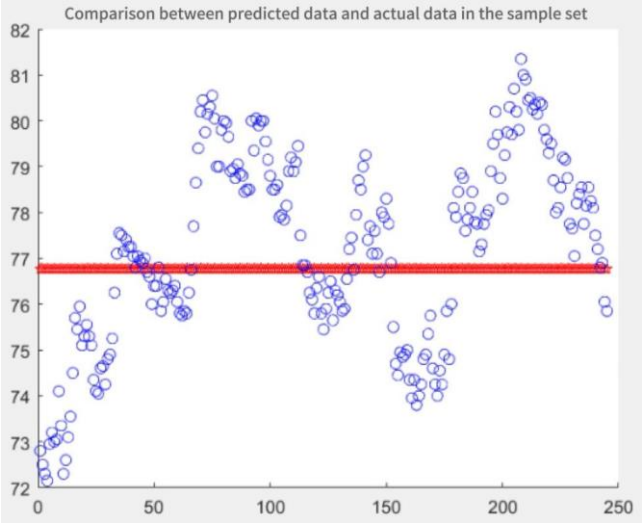

Figure 11. Comparison between predicted data and actual data in the test sample set

From Figures 8 and 9, it can be seen that there is a relatively large error between the calculated output and the actual output at the beginning, and the last set of data has the smallest error. The weight matrices of the trained network's hidden layer and output layer are W1 and W2, respectively. After inputting the test sample set into the network, the error is 2.2473, which is slightly larger than the error of the training sample set.

From Figure 10, it can be seen that for the 492 sets of

training samples, the overall fit between the calculated output and the actual sample output is relatively good. Although there were several sets of data with significant fluctuations in the middle, the algorithm adjusted the weights in a timely manner. It can be seen that the last few sets of sample data fit relatively well, indicating that the error eventually converged.

From Figure 11, it can be seen that the error between the calculated output and the actual sample output obtained from 246 sets of test samples is larger than that of the training sample set, indicating that the designed network has poor generalization ability. After analysis, one reason is that there are too few training samples, and the other is that the prediction model is too simple, and the closing price is also related to other data.

4. Summary

This system first collects data on stocks during a certain period of time, including various data that can reflect the nature of stocks, such as stock indices, opening prices, closing prices, trading volumes, and some statistical data that can be used for macroeconomic analysis.

Subsequently, this article applies the BP neural network algorithm in MATLAB to simulate and predict the future trend of stocks, in order to judge the quality of stocks and the buying and selling strategies. Due to the fact that the stock market is a highly complex nonlinear dynamic system, the BP neural network method can achieve a highly nonlinear mapping from R space to R_n space using only samples. It has good adaptability to atypical data and obvious superiority in dealing with missing values and nonlinear problems. However, due to the complexity of stock data, a single prediction method is limited. How to use multiple methods to comprehensively make qualitative and quantitative predictions is also the focus of future research on stock prediction.

References

[1] Hu Jie, Zeng Xiangjin BP neural network stock price prediction based on LM algorithm [J] Technology Entrepreneurship Monthly, 2007, 20 (2): 183-183

[2] Zhang Jigang An improved method for stock price prediction based on BP neural network [J] Journal of Changjiang University A (Natural Science Edition), 2007

[3] Zhang Caifeng PSO-BP neural network stock price prediction [J] Business Manager, 2016 (9)

[4] Zhang Soaring Research on Domestic and Foreign Stock Market Prediction Based on BP Neural Network Stock Index Prediction System [D] Hunan University, 2017

[5] Wu Haihua, Ma Yuan, etc BP neural network stock index prediction model [J] Journal of the Communist Party of China Qingdao Municipal Party School and Qingdao Administrative College, 2003

[6] Yang Xiaoyan The Application of Neural Networks in Stock Market Prediction [D] North China University, 2008

[7] Fu Shifeng, Zhao Li, Cai Wenjun Research on the Application of BP Neural Network in Stock Market Prediction [J] Computer Knowledge and Technology: Academic Edition, 2021

[8] Bai Danjin Xin Sun Fangfang Establishment of Stock Price Prediction Model Using BP Neural Network [J] Technology Entrepreneurship Monthly, 2006

[9] Yang Jie, Jia Shuwen, Xu Siheng, Che Lujun Research on Stock Price Prediction Based on BP Neural Network [C]//2020 Wanzhi Science Development Forum 0

[10] Yu Haishu, Cai Jihua, Xia Hong The Application of ARIMA Model in Stock Price Prediction [J] Economist, 2015 (11): 2

[11] Dong Bolun, Xu Dongyu Prediction and Analysis of Agricultural Product Stock Prices Based on ARIMA Model [J] Modern Business, 2015 (3): 3

[12] Chen Huan, Chen Peng Using ARIMA model to predict stock prices - taking Laibao High tech (002106) as an example [J] Business Situation, 2012 (8): 1

[13] Zhao Kangyin, Xue Yanan TCL Group Stock Prediction and Evaluation Based on ARIMA Model [J] two thousand and nineteen

[14] Shi Yang Analysis of the Linkage between Stock Prices and the Renminbi Exchange Rate: Based on the Copula ARIMA Model [J] Journal of Shanxi University of Finance and Economics, 2019

[15] Han Junwei Analysis and Prediction of Shanghai Stock Exchange 50 Index Based on ARIMA Model [J] Fujian Quality Management, 2019 (24)