

A Short-term Prediction about the Stock Price of China's Pharmaceutical Manufacturing Industry based on Back Propagation Neural Network

Xiang Li

College of Finance and Statistics, Hunan University, Changsha 410006, China

1119044359@qq.com

Abstract. This paper applies Back Propagation neural network to make a short-term prediction about the average stock price of pharmaceutical manufacturing industry in the next 30 days after October 23rd, 2018. The results show that the price will generally increase; the frequency of price fluctuation will be relatively high; price volatility will be relatively low.

Keywords: Time series, pharmaceutical manufacturing industry stock price, daily data, short-term prediction.

1. Introduction

Against the context that China's population aging becomes more serious, the government pays more attention to healthcare field, and the proportion of medical expenditure to GDP grows year by year, the pharmaceutical manufacturing industry has increasingly become an important industry related to the national prosperity and people's livelihood. At the same time, the increasing demands and great potential of this industry also attract a large number of investors. Therefore, focusing on the development of this industry, analyzing and forecasting the stock price of listed companies in this industry will not only help listed companies adjust their strategies and improve management, but also help investors to make rational choices and expand social wealth and welfare.

In China's financial market, most investors are retail ones who mainly invest in the short-term, so the short-term trend and fluctuation of the stock price are particularly crucial for investors. Therefore, based on the historical average stock price of the pharmaceutical manufacturing industry, this paper applies the BP neural network model to predict the short-term stock price in the next 30 days.

Besides, the significance of this study is manifested in the following two perspectives. First, the theoretical significance of this study is to expand the research on the stock price of China's pharmaceutical manufacturing industry. The stock price fluctuates from time to time, which disturbs the normal production and consumption of enterprises and residents. The stock price of pharmaceutical manufacturing industry has gradually attracted the attention of the society and some scholars. This paper predicts the stock price and compares the effects of Back Propagation neural network (BP neural network) in prediction, which will make a contribution to expand the research content of pharmaceutical manufacturing economy. Second, the practical significance is that by forecasting the stock price of the pharmaceutical manufacturing industry, this study will provide management countermeasures against stock price fluctuations for this industry, enhance its ability to resist the risk of stock price fluctuations, and promote the healthy and sustainable development of this industry. The research methodology in this paper is in line with the actual situation, which can be directly and generally applied in practice.

2. Literature Review

In the financial time series, ARMA model and BP neural network are usually used to simulate the trend of financial price. In western developed countries, with an early origin of the stock exchange market, the theoretical research and quantitative analysis of financial time series are more systematic. In terms of research methods, price prediction research has gone through a long process from qualitative analysis to quantitative analysis. The qualitative prediction, based on the subjective judgment of human beings to forecast the future trend of things, is ambiguous and difficult to make

accurate estimations, which is generally feasible to predict the event without historical statistics or the one with turning points. However, the quantitative prediction uses mathematical models to calculate the future value of things. There are three main quantitative research methods for price prediction at home and abroad: simple linear prediction, general nonlinear prediction, and modern artificial intelligence prediction. Simple linear prediction methods include the AR model and its special cases--the VAR model, ARMA model, and ARIMA model. The general nonlinear prediction methods contain ARCH model, GARCH model and EGARCH index model, which can predict the nonlinear relations and non-stationary sequences, but the prediction result is unsatisfactory. Artificial intelligence prediction methods mainly include artificial neural network, support vector machine and wavelet analysis, of which the artificial neural network and the support vector machine are more complicated in parameter setting, while the wavelet analysis must be used together with other methods. Therefore, this paper chooses to use the artificial neural network method to make a short-term prediction about the stock price of China's pharmaceutical manufacturing industry. In addition, the artificial neural networks consist of more than 200 kinds of networks, among which BP neural network is the most widely used, so BP neural network is adopted in this paper to make a short-term prediction about the stock price of China's pharmaceutical manufacturing industry.

In this paper, the time series of the average closing price of pharmaceutical manufacturing industry are taken as the research object, and a simple and feasible prediction model is proposed to make a short-term prediction about the stock price of the pharmaceutical manufacturing industry. Apart from that, by simulating and analyzing the previous data, the trend of the price in the next 30 days after October 23rd, 2018 is also predicted.

3. Research Methods

In 1943, the neurophysiologist Warren Mc Culloch along with the mathematician Walter Pitts proposed a neural network model. Inspired by graph theory in computer, they reckoned that the logic function of human neurons is the same as that of computer binary code--1 and 0, and then proposed a network structure and mathematical description about neuron, which marks the beginning of artificial neural network era and provides new ways for people to understand and create artificial intelligence. In 1957, Frank Rosenblatt, an experimental psychologist at Cornell University, implemented a neural network model by experimental simulation on an IBM computer, named "Perceptron", according to the learning mechanism understood by people at that time--the efficiency of the cell junction can be increased when one neuron repeatedly activates another. In the early 1980s, Hopfield, a professor of biophysics at California Institute of Technology, proposed a new neural network which provides an approximation algorithm for combinatorial optimization problems.

In 1986, the Back Propagation neural network, also known as backward propagation of errors, was first proposed by Professor Rumelhart and Professor Mc Clland in the book *Parallel Distributed Processing*. It adopts backward propagation algorithm to update the coefficient in each layer of the neural network in accordance with the error between network output value and the actual value. The BP neural network algorithm covers a wide range, and is available to most of the continuous functions. Up to now, the BP neural network is still a popular topic in the relevant research of neural networks. In 2006, Professor Hinton, from the University of Toronto, first proposed the concept of deep learning. In this method, an unsupervised learning strategy, also called layer-wise training, is adopted to extract more abstract and representative features from primary data and effectively train a neural network, called deep belief network. Later, Professor Hinton and other scholars proposed the Dropout method for reducing overfitting in neural network, and proved that layer-wise training can effectively improve the generalization ability of the deep belief network [1]. In 2014, Shinsuke Kimuraa, Kunikazu Kobayashib, and Masanao Obayashia constructed time series forecasting using a deep belief network with restricted Boltzmann machines [2]. In 2016, Matthew Dixon, Diego Klabjan, and Jin Hoon Bang implemented a prediction about the pharmaceutical manufacturing market under the guidance of deep learning neural network [3]. In the next sections, a brief introduction to BP neural network will be

given, and a short-term prediction about the stock price of pharmaceutical manufacturing industry based on BP neural network will be presented.

3.1 Theoretical Model of BP Neural Network

A BP neural network consists of multiple network layers, including an input layer, several hidden layers, and an output layer. Besides, its learning process is made up of feed-forward propagation and backward propagation. In feed-forward propagation, information moves from the input layer, through the hidden layer, and to the output layer. The state of the neurons in each layer only affects the state of the neuron in the next layer. If the expected output is not generated in the output layer, the process will be changed to backward propagation and the error signal will go back through the original neural network. In the process of returning, the connection weights and thresholds of the neurons in each layer will be modified one by one. This process is continuously iterated until the error signal can satisfy the criterion.

A typical BP neural network with three layers is illustrated in Figure 1.

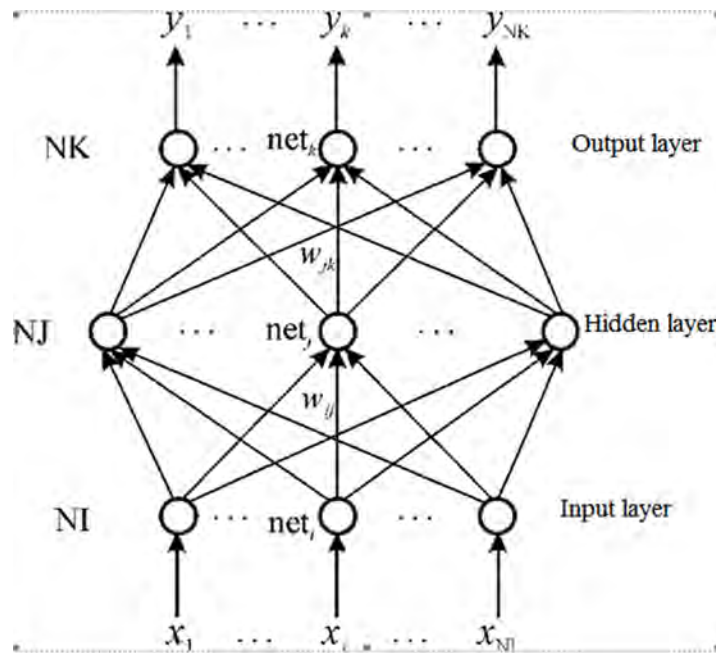


Figure 1. Typical BP neural network with three layers

Assuming the number of the neuron in the input layer, hidden layer and the output layer of the BP network is NI, NJ and NK respectively, the input of the neuron j in the hidden layer is defined as:

$$net_j = \sum_{i=1}^{NI} w_{ij} o_i ; j=1, 2, \dots, NJ \quad (1)$$

where w_{ij} is the weight in the connection of neuron i in the input layer to neuron j in the hidden layer, while o_i is the output of neuron i in the input layer.

The input of neuron k in the output layer is defined as:

$$net_k = \sum_{j=1}^{NJ} w_{jk} o_j ; k=1, 2, \dots, NK \quad (2)$$

where w_{jk} is the weight in the connection of neuron j in the hidden layer to neuron k in the output layer, while o_j is the output of neuron j in the hidden layer.

The outputs in the input layer, the hidden layer and the output layer are respectively calculated:

$$o_i = net_i = x_i \quad (3)$$

$$o_j = f_j(net_j, \theta_j) = \frac{1}{1 + e^{-(net_j - \theta_j)}} \quad (4)$$

$$y_k = o_k = f_k(net_k, \theta_k) = \frac{1}{1 + e^{-(net_k - \theta_k)}} \quad (5)$$

where θ_j is the threshold of neuron j in the hidden layer, while θ_k is the threshold of neuron k in the output layer.

Assuming that the number of training samples is p, the input vectors are x^1, x^2, \dots, x^p , the output vectors are y^1, y^2, \dots, y^p , and the corresponding expected output (sample) vectors are t^1, t^2, \dots, t^p , the mean square error of example p can be defined as:

$$E^p = \frac{1}{2} \sum_{k=1}^{NK} (t_k^p - y_k^p)^2 \quad (6)$$

where t_k^p and y_k^p respectively denote the expected and actual output of the output neuron k of the example p.

The weight in the output layer can be updated:

$$\Delta w_{jk}(n+1) = \eta \delta_k^p o_j^p + \alpha \Delta w_{jk}(n) \quad (7)$$

$$\delta_k^p = (t_k^p - y_k^p) f'_k(net_k^p) \quad (8)$$

where η denotes the learning rate, while α is a variable inertia term (Momentum).

The weight in the hidden layer can be updated:

$$\Delta w_{ij}(n+1) = \eta \delta_j^p o_i^p + \alpha \Delta w_{ij}(n) \quad (9)$$

$$\delta_j^p = f'_j(net_j^p) \sum_{k=1}^{NK} \delta_k^p w_{jk} \quad (10)$$

The threshold is updated in the same method as the weight.

Determination of the number of the unit in the hidden layer:

A common method of determining the optimal number of hidden layer units is the trial and error method, which refers to firstly setting up fewer hidden layer units to train the network, then gradually increasing the number of hidden layer units until the minimum error in the network is obtained. In the trial and error method, some empirical formulas can be taken as a reference:

$$m = \sqrt{n+1} + a$$

where m, n, and 1 respectively denote the number of the units in the hidden layer, the input layer, and the output layer, and a is a constant term between 1 and 12. In this paper, this formula is adopted as the reference to determine the approximate range of the number of the units in the hidden layer, and the specific number is determined in the method of trial and error.

3.2 The Construction of BP Neural Network

In this paper, the research object is China's pharmaceutical manufacturing industry stocks, and the research data are the 6833 daily transaction records from January 14th, 1991 to October 23rd, 2018, of which the previous 4783 transaction records are taken as the training group, the next 1025 records are selected as the test group, and the last 1025 records are chosen as the verification group. A three-layer neural network model is constructed with the average price AVPRICE as the input variable and the predicted price EXPRICE as the output variable. As Figure 2 shows, the number of neurons in the input layer is 1, the number of the hidden layers is 10, and the number of output layers is 1. The standard Sigmoid function is adopted in the output layer and the hidden layers:

$$f_j(net_j, \theta_j) = \frac{1}{1 + e^{-(net_j - \theta_j)}}$$

Besides, Figure 3 below illustrates the training process of neural network.

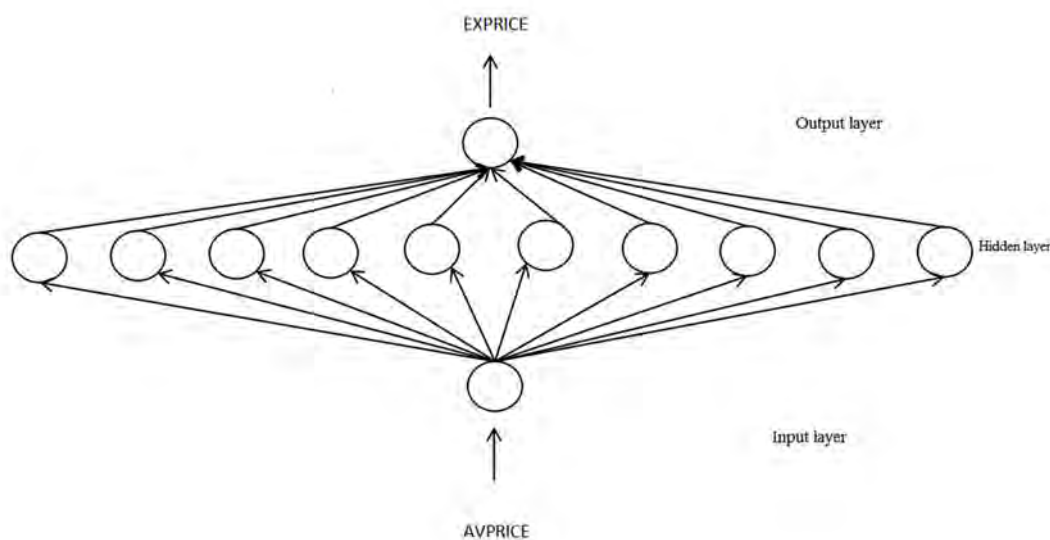


Figure 2. the structure of BP neural network

The specific steps are as follows:

Step 1: Initialize the training by choosing the weight and the threshold at random, determining the learning rate and the momentum, and setting the learning error criterion.

Step 2: Select the average closing price of the previous 4783 stocks as the training samples.

Step 3: Conduct feed-forward propagation by calculating the output of each layer according to equations (1) to (5).

Step 4: Conduct backward propagation by calculate the error of each layer with the equations (8) and (10).

Step 5: Modify the weights and thresholds in accordance with equations (7) and (9).

Step 6: Judge whether all the samples have been trained. If not, return to Step 3 to train the next sample.

Step 7: Calculate the total performance indicator. Calculate the error function of all samples:

$$E_{all} = \sum_{p=1}^p E^p$$

to determine whether it meets the criterion: $E_{all} < \varepsilon$. If it is satisfied, the training ends; otherwise, return to Step 3.

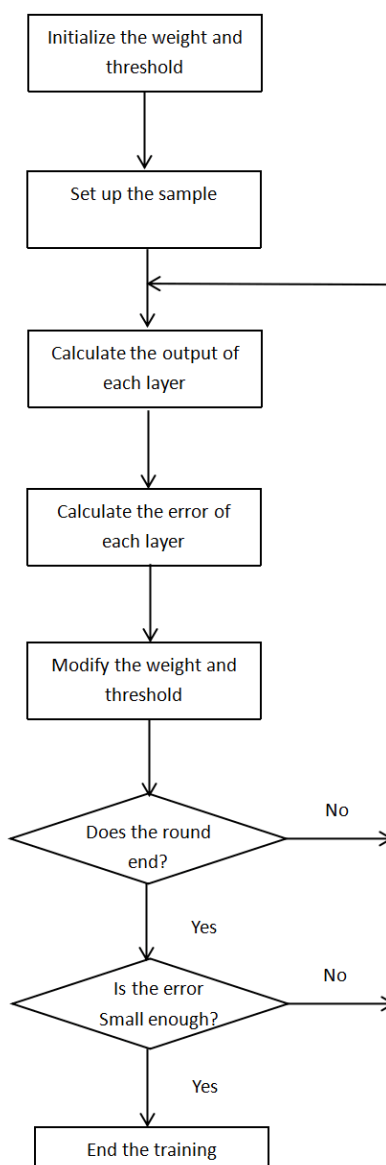


Figure 3. the training process of BP neural network

4. The Source of Data and the Selection of Samples

In China, there are 215 pharmaceutical manufacturing industry stocks, including Yunnan Baiyao, Renhe Pharmacy, Dong'e Ejiao, Tongrentang, Jiangzhong Pharmacy, and so on.

The selection criteria are as follows: first, the listed companies should have the capacity of sustainable management; second, the stock price fluctuates in the normal range, without any continuous abnormal value.

Among the 215 stocks, the price of Shenqi Pharmaceutical stock was significantly higher than the normal range at the initial stage, and only lasted for a short period, which indicates that its stock price during this period obviously deviated from its actual value, and was greatly affected by accidental factors. From this perspective, the Shenqi Pharmaceutical stock is not representative and reliable, and cannot be selected as the sample. Therefore, in this paper, only the closing prices of 214

pharmaceutical manufacturing industry stocks from January 14th, 1991 to October 23rd, 2018 are selected as the samples.

5. Empirical Analysis

5.1 Descriptive Statistic

First, the average closing price of the selected 214 stocks is calculated, and the average price of the pharmaceutical manufacturing industry stocks is taken as the research object. The descriptive statistics are shown in Table 1.

Table 1. the descriptive statistic of the main variable

Variable	Symbol	Samples	std	max	min	mean
The average closing price of China's pharmaceutical manufacturing industry stocks	AVPRICE	6833	7.10	47.23	3.15	15.48

5.2 Prediction Results

The short-term prediction results of the average closing price of China's pharmaceutical manufacturing industry stocks are estimated based on BP neural network. The goodness of fit of the model is shown in Table 2. The absolute and relative errors are shown in Table 3.

Table 2. Goodness of fit

	Goodness of fit
Goodness of fit of training	0.99852
Goodness of fit of modification	0.99837
Goodness of fit of text	0.9977
Overall goodness of fit	0.99837

Table 3. the prediction effect of BP neural network

Error Types	
Average absolute error	Average relative error
-0.000708078	0.00036455

As Table 2 shows, the goodness of fit of training is 0.99852, the goodness of fit of modification is 0.99837, the goodness of fit of test is 0.9977, and the overall goodness of fit is 0.99837, which indicates that the BP neural network fits relatively well. In Table 3, the average absolute error is -0.000708078 and the average relative error is 0.00036455, which shows that there is no overfitting performance in the short-term prediction of BP neural network, and the overall prediction effect is well. What's more, the results of the short-term prediction are shown in Table 4.

In Figure 4 below, the actual value and predicted value of the average stock price of the pharmaceutical manufacturing industry from January 14th, 1991 to October 23rd, 2018 are described. The figure shows the following trends of the stock price in the next 30 days after October 23rd, 2018:

- (1) The stock price will generally increase;
- (2) The stock price will fluctuate every day and the fluctuation frequency will be high;
- (3) The average stock price will increase from 16.25 yuan to 16.59 yuan, with an increase of 2.09%, and price volatility will be relatively low.

Table 4. the short-term prediction results of BP neural network

Sequence Number	Predicted Value	Sequence Number	Predicted Value
1	16.247751	16	16.427412
2	16.25984	17	16.439118
3	16.272028	18	16.450789
4	16.284187	19	16.462423
5	16.296312	20	16.474021
6	16.308404	21	16.485583
7	16.320461	22	16.497109
8	16.332484	23	16.508598
9	16.344473	24	16.52005
10	16.356426	25	16.531465
11	16.368345	26	16.542843
12	16.380229	27	16.554185
13	16.392078	28	16.565489
14	16.403892	29	16.576756
15	16.41567	30	16.587986

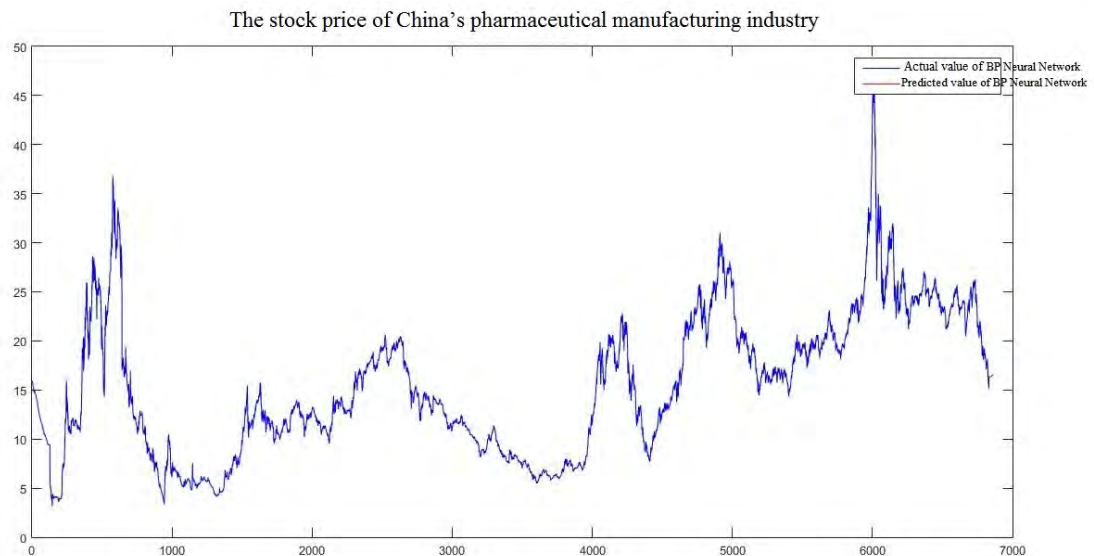


Figure 4. the actual value and the predicted value of the pharmaceutical manufacturing industry average stock price

6. Conclusion

In this paper, in the view of BP neural network, a short-term prediction about the daily closing price of pharmaceutical manufacturing industry stocks from January 14th, 1991 to October 23rd, 2018 is conducted. In terms of goodness of fit, absolute error and relative error, the prediction effect of this model is well. The prediction results show that in the next 30 days, the trends of the daily closing price of China's pharmaceutical manufacturing industry stocks can be concluded as follows: the price will generally increase; the frequency of price fluctuation will be relatively high; price volatility will be relatively low.

References

- [1]. Hinton G E, Salakhutdinov R. Reducing the Dimensionality of Data with Neural Networks [J]. Science, 2006, 313(5786): 504-507.

- [2]. Kuremoto T, Kimura S, Kobayashi K, et al. Time series forecasting using a deep belief network with restricted Boltzmann machines[J]. *Neurocomputing*, 2014, 137(15):47-56.
- [3]. Dixon M F, Klabjan D, Bang J H. Classification-Based Financial Markets Prediction Using Deep Neural Networks [J]. *Social Science Electronic Publishing*, 2016, 3: 28-37.