Journal of Computers Vol. 30 No. 3, 2019, pp. 260-268
doi:10.3966/199115992019063003021

Stock Selection in Investment Management of Commercial Stock Market: Prediction by Data Mining

Qun Wang[1*], Xiaoyu Hu[2]

[1] Tongling University, Tongling, Anhui 244061, China
qunw_wang@126.com

[2] Jinan University, Guangzhou, Guangdong 510632, China
xiaoyh_xy@yeah.net

Received 17 April 2019; Revised 29 April 2019; Accepted 8 May 2019

Abstract. Stock forecasting is an important part of stock market investment management. In this study, stock was predicted using data mining method, and back propagation neural network (BPNN) was taken as the basis. Genetic algorithm was used for optimization to obtain the improved BPNN algorithm, and then it was applied to stock prediction. Instance analysis was performed taking a stock as an example. The forecasting results of BPNN, particle swarm optimization (PSO) improved BPNN and the improved method were compared. The results showed that the forecasting results of the improved BPNN method were basically consistent with the actual values, with an average error of 0.8%, much lower than 7.19% of the BPNN algorithm and 4.17% of the PSO-BPNN algorithm, suggesting that the improved BPNN algorithm was valid. Then four stocks were predicted to understand the future development of those stocks and help select proper stock for investment. It suggested that the improved BPNN algorithm could provide a reliable basis for stock selection. This study provides some theoretical support for the further application of data mining in stock forecasting, which is helpful for investors to make correct stock choices, improve returns and avoid risks. Moreover, it also contributes to stabilizing the stock market and promoting economic development.

Keywords: data mining, genetic algorithms, neural network, stock prediction, stock selection

1 Introduction

With the development of economy, the stock market has become an important part of the financial market, which is favored by more and more investors and becomes a new investment choice. Stock is a high-risk and high-yield investment mode, and the volatility of the stock market will have an impact on the financial market. Therefore, stock forecasting is an important and challenging task [1]. The prediction of stock price is not only of great significance to investors in investment management, but also conducive to the healthy development of our economy. The change of stock market is influenced by many factors [2]. Moreover stock market has characteristics of non-linearity and high noise. It is difficult to predict stock effectively by traditional point graph method and line graph method [3]. With the development of computer technology, data mining method has been widely used in stock prediction [4]. Huang et al. [5] improved the quality of input features using the sub-mode coordinate algorithm, then predicted stock using the improved neural network model, and proved the effectiveness of the method through an example analysis. Sadaei et al. [6] optimized the method of fuzzy time training set by using imperialist competitive algorithm, verified the method by example analysis, and found that it had higher accuracy than other methods. Wang et al. [7] selected the input variables of support vector machine using principal components, then searched for the best parameters through brainstorming optimization, and proved the effectiveness of the hybrid prediction algorithm through actual case analysis. Bini et al. [8] analyzed clustering and regression in data mining and then predicted stock market using multiple regression

* Corresponding Author

algorithm to help investors choose companies with higher profits. Neural network is one of the methods of data mining. It can be found that neural network has excellent performance in prediction and has become one of the hotspots of international research. Back propagation neural network (BPNN) with good fault tolerance and generalization is the most widely used neural network at present. In order to make BPNN perform better in various applications, the optimization of BPNN using particle swarm optimization (PSO), adaptive learning rate, etc., has also been deeply studied. In this paper, genetic algorithm (GA) was used to optimize BPNN, and then its performance in stock forecasting was analyzed. It was found that this method could basically achieve accurate stock forecasting, which can provide some references for stock selection. The research results proved that the optimized BPNN method designed in this paper had good performance in stock forecasting and selection, which is worth application and promotion.

2 Data Mining and Stock Forecasting

Stock forecasting is an important part of stock market investment management. Investors need to have an understanding of the changing rules of stocks and make relatively accurate predictions, in investment management before selecting the appropriate stocks for investment. Stock forecasting is based on a large number of stock trading data, and with the development of the stock market, its cumulative data is also in the explosive growth [9]. Traditional forecasting methods has not been able to deal with these massive data effectively. Therefore, investors urgently need effective stock forecasting methods. Data mining refers to extracting useful information from a large number of random data containing noise to provide support for decision-making, including artificial intelligence, machine learning and other technologies. In the environment of big data, data mining has been widely used in many fields because of its advantages in big data processing [10]. It also has good performance in stock forecasting, such as neural network [11], clustering algorithm, multiple regression [12]. Analyzing and processing massive amount of historical stock information using data mining method can obtain the law of stock changes and help make the best stock selection in the stock market investment management. This study mainly explored the application of BPNN method in stock forecasting.

3 BPNN

BPNN is one of the methods of data mining. It has good non-linear fitting ability and is very suitable for stock prediction. It consists of input layer, hidden layer and output layer (Fig. 1). The learning process includes forward propagation of signals and back propagation of errors.

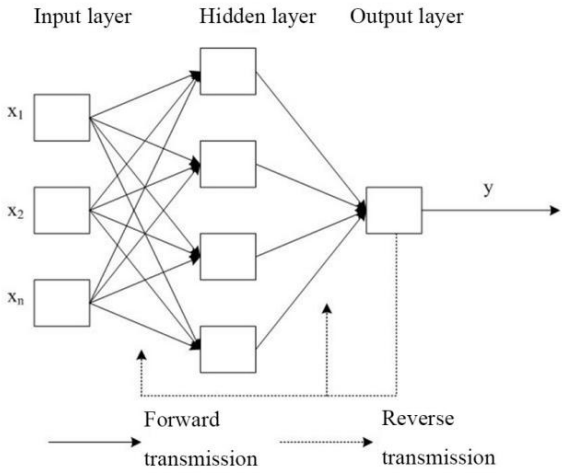

Fig. 1. BPNN

Suppose the input layer as n, the input vector as $X=(x_1,x_2,\cdots,x_n)^T$, the hidden layer node as h, the output vector as $O=(o_1,o_2,\cdots,o_h)^T$, the output layer as m, the output vector as $Y=(y_1,y_2,\cdots,y_m)^T$, and

the expected output vector as $D=(d_1,d_2,\cdots,d_m)^T$. The connection weight matrix between different layers were expressed as V and W, and learning rate was expressed as η. Then the input and output of the j-th neuron in the output layer can be expressed as $y_k=f\left(\sum_{j=1}^{h}w_{jk}y_jk\right),k=1,2,\cdots,m$, where $f(x)$ represents the excitation function.

When the actual output does not coincide with the expected output, the output error is expressed as E, $E=\frac{1}{2}(D-Y)^2=\frac{1}{2}\sum_{k=1}^{m}(d_k-y_k)^2$. The error can be reduced by adjusting the weight of each layer:

$$\begin{aligned}\Delta v_{ij}&=-\eta\frac{\partial E}{\partial v_{ij}},i=1,2,\cdots,n\\ \Delta w_{ij}&=-\eta\frac{\partial E}{\partial w_{jk}},j=1,2,\cdots,h\end{aligned}. \tag{1}$$

According to $y_k=f\left(\sum_{j=1}^{h}w_{jk}y_jk\right),k=1,2,\cdots,m$, we have:

$$E=\frac{1}{2}(D-Y)^2=\frac{1}{2}\sum_{k=1}^{m}\left(d_k-f\left(\sum_{j-1}^{h}w_{jk}y_j\right)\right)^2. \tag{2}$$

Then the weight adjustment of each layer can be expressed as:

$$\begin{aligned}\Delta w_{jk}&=\eta(d_k-y_k)f\left(\sum_{j=1}^{h}w_{jk}y_jk\right)\\ \Delta v_{ij}&=\eta(d_k-y_k)f\left(\sum_{j=1}^{h}w_{jk}y_jk\right)w_{jk}\end{aligned}. \tag{3}$$

The algorithm flow of BPNN is shown in Fig. 2.

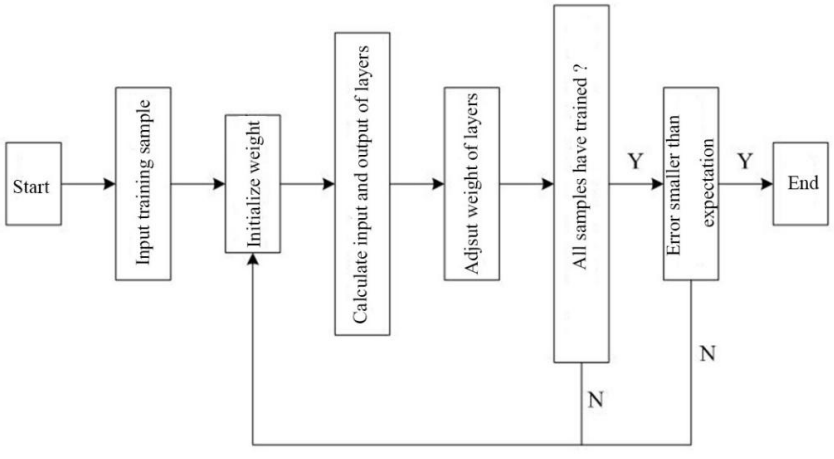

Fig. 2. The algorithm flow of BPNN

Firstly, the weights are initialized to determine the learning rate, excitation function and target error. Then, forward propagation and error back propagation are carried out for every sample which is input to network, and the weights are constantly adjusted to make the actual output of the network closer to the expected output.

The results showed that the BPNN model could effectively take into account various non-linear factors affecting the stock price and has relatively high efficiency and accuracy in stock market prediction.

4 Stock Forecasting Method Based on Optimized BPNN Algorithms

Although BPNN has excellent self-adaptability and self-learning ability, there are still some problems in the practical application of the algorithm, such as slow speed and easy to fall into local minimum. In this study, BPNN was improved using Genetic algorithm (GA).

GA is a method of searching global optimal solution which simulates the evolution law of the survival of the fittest in the biological world, as shown in Fig. 3.

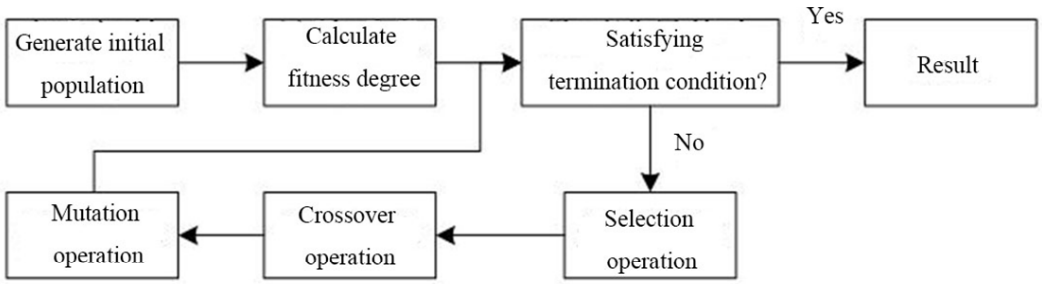

Fig. 3. Genetic algorithm

BPNN was optimized by GA and applied to stock forecasting, and the specific procedures are as follows.

(1) A group of initial population was generated randomly and coded by real number. Each individual in the population represented an initial weight of BPNN, and each gene value represented the connection weight of BPNN.

(2) Output error E of BPNN was calculated to determine the fitness function: $f(x)=\frac{1}{E+1}$.

(3) Individuals were arranged by fitness proportion method, and individuals with larger fitness values were inherited to the next generation.

(4) Cross-over operation was made. The cross-over probability is between 0.4 and 1 generally, 0.8 in this study, and individuals in the population were randomly selected for exchange.

(5) Mutation operation was made. The mutation probability is between 0.01 and 0.1 generally, and uniform mutation was adopted in this study, i.e., random number, to form a new generation of population.

(6) Steps (2)-(5) were repeated until training objectives were met.

The stock forecasting results were obtained by substituting the optimized weights into the neural network.

5 Case Study

5.1 Algorithm Reliability Analysis

A stock is randomly selected from RESSET database. The data including opening price, highest price, lowest price and closing price from January 1, 2018 to December 14, 2018 were selected as training samples. The stock data from December 3, 2018 to December 27, 2018 were selected as testing samples. In the design of BPNN, a three-layer network which contained a hidden layer was selected; the number of nodes in the input layer was 6, the number of nodes in the output layer was 1, and the number of nodes in the hidden layer was set as 11 after multiple times of test.

Some experimental data are shown in Table 1 and Table 2.

In order to validate the effectiveness of BPNN, PSO improved BPNN and the improved BPNN proposed in this study were used for stock forecasting. The prediction results are shown in Fig. 4 to Fig. 7.

It was found from Fig. 4 to Fig.7 that the predicted value of the improved BPNN was closer to the actual stock value index, and the two curves basically coincided. When BPNN and PSO-BPNN was used, there were some errors and large fluctuations. The average errors of the three methods are shown in Table 3.

Table 1. Data of the training samples

Date	Opening price	The highest price	The lowest price	Closing price
2018-01-02	7.34	7.42	7.29	7.35
2018-01-03	7.39	7.54	7.3	7.45
2018-01-04	7.4	7.44	7.31	7.35
2018-01-05	7.37	7.39	7.3	7.32
2018-01-08	7.65	7.65	7.31	7.42
2018-01-09	7.37	7.38	7.2	7.25
2018-01-10	7.23	7.23	7.1	7.16
2018-01-11	7.13	7.24	7.12	7.21
2018-01-12	7.18	7.22	7.16	7.17
2018-01-15	7.17	7.17	6.8	6.9
2018-01-16	6.88	6.97	6.85	6.9
2018-01-17	6.84	6.92	6.71	6.86
2018-01-18	6.82	6.98	6.82	6.88
2018-01-19	6.9	6.97	6.84	6.86
2018-01-22	6.86	7.55	6.75	7.55
2018-01-23	7.65	7.72	7.33	7.45
2018-01-24	7.34	7.64	7.25	7.48
2018-01-25	7.37	7.54	7.26	7.28
2018-01-26	7.3	7.52	7.26	7.36
2018-01-29	7.3	7.43	7.22	7.27
2018-01-30	7.26	7.43	7.24	7.42
2018-01-31	7.38	7.42	7.02	7.05

Table 2. Data of the test samples

Date	Opening price	The highest price	The lowest price	Closing price
2018-12-03	3.85	3.99	3.82	3.98
2018-12-04	4	4.07	3.95	4.01
2018-12-05	3.95	4.13	3.9	4.07
2018-12-06	4.01	4.11	3.96	3.96
2018-12-07	4.02	4.05	3.93	3.93
2018-12-10	3.91	3.92	3.81	3.86
2018-12-11	3.87	3.9	3.84	3.89
2018-12-12	3.9	3.97	3.86	3.88
2018-12-13	3.94	3.94	3.82	3.87
2018-12-14	3.85	3.86	3.72	3.73
2018-12-17	3.74	3.75	3.66	3.72
2018-12-18	3.69	3.73	3.62	3.66
2018-12-19	3.68	3.69	3.58	3.58
2018-12-20	3.58	3.63	3.55	3.61
2018-12-21	3.58	3.6	3.47	3.53
2018-12-24	3.55	3.6	3.51	3.58
2018-12-25	3.54	3.55	3.38	3.5
2018-12-26	3.52	3.53	3.45	3.47
2018-12-27	3.55	3.55	3.36	3.36
2018-12-28	3.36	3.44	3.36	3.42

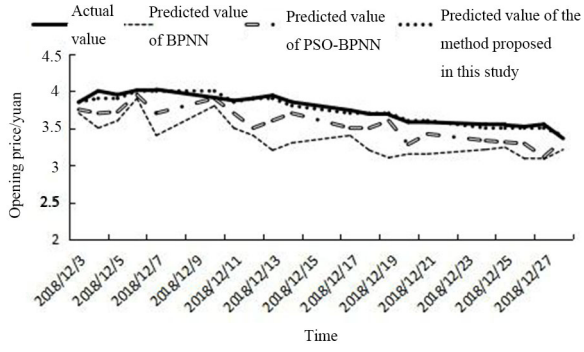

Fig. 4. Prediction results of opening price

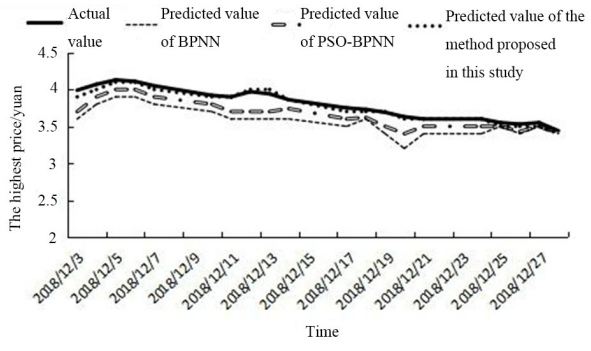

Fig. 5. Prediction results of the highest price

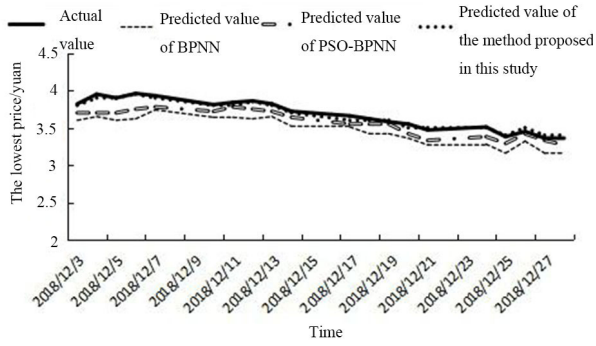

Fig. 6. Prediction results of the lowest price

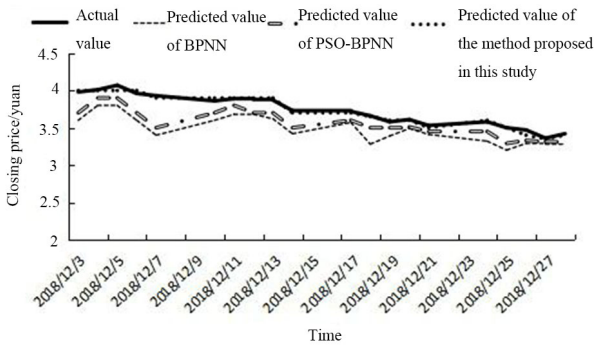

Fig. 7. Prediction results of closing price

Table 3. Comparison of average error

	BPNN	PSO-BPNN	The improved BPNN
Opening price	10.59%	5.58%	0.92%
The highest price	5.99%	3.80%	0.85%
The lowest price	5.69%	2.98%	0.76%
Closing price	6.49%	4.30%	0.68%
Average error	7.19%	4.17%	0.80%

According to Table 3, it was found that the error of the improved BPNN was significnatly smaller than that of BPNN and PSO-BPNN. The average error of BPNN was 7.19% and that of PSO-BPNN was 4.17%, while the average error of the improved BPNN was 0.8%, indicating the effectiveness of improving BPNN with GA and the reliability of the improved BPNN in the prediction of stock. The improved BNPP can make an accurate prediction on stock changes and provide investors with reliable basis in stock selection.

5.2 Application of Algorithm in Stock Selection

Four stocks were randomly selected from the RESSET database, and some data of the stocks are shown in Table 4.

BPNN was training using the data of the four stocks, and then the trend of the stocks in the next 15 days was predicted. The predicted results are shown in Fig. 8 to Fig. 11.

It was found from Fig. 8 to Fig. 11 that the changes of stock A and B were stable, and there was almost no fluctuation, indicating that these two stocks will not bring investors greater returns or cause greater losses in the short term. There was a small fluctuation in stock C, and there was an upward trend in the latter period, showing that this stock could be invested continuously. Stock D showed large fluctuations, indicating that stock D had high risks and was not suitable for investment.

Table 4. Training data

Stock	Date	Opening price	The highest price	The lowest price	Closing price
A	2018-01-02	13.35	13.93	13.32	13.7
	2018-01-03	13.73	13.86	13.2	13.33
	2018-01-04	13.32	13.37	13.13	13.25
	2018-01-05	13.21	13.35	13.15	13.3
	2018-01-08	13.25	13.29	12.86	12.96
	2018-01-09	12.96	13.2	12.92	13.08
	2018-01-10	13.04	13.49	12.92	13.47
	2018-01-11	13.41	13.59	13.27	13.4
	2018-01-12	13.45	13.68	13.41	13.55
	2018-01-15	13.51	14.33	13.5	14.2
B	2018-01-02	8.4	8.46	8.3	8.41
	2018-01-03	8.41	8.93	8.41	8.67
	2018-01-04	8.59	8.78	8.5	8.72
	2018-01-05	8.72	8.74	8.62	8.63
	2018-01-08	8.64	8.68	8.37	8.43
	2018-01-09	8.4	8.69	8.31	8.54
	2018-01-10	8.56	8.59	8.26	8.35
	2018-01-11	8.33	8.34	8.17	8.29
	2018-01-12	8.25	8.34	8.18	8.2
	2018-01-15	8.17	8.17	7.6	7.65
C	2018-01-02	35.01	35.27	34.88	35.12
	2018-01-03	35.16	35.49	34.8	35.36
	2018-01-04	35.36	35.6	35.02	35.37
	2018-01-05	35.37	35.37	34.55	35
	2018-01-08	34.98	34.98	34.26	34.35
	2018-01-09	34.36	34.88	34.3	34.63
	2018-01-10	34.47	35.18	34.36	35.09
	2018-01-11	35	35.65	34.8	35.61
	2018-01-12	35.53	35.87	35.1	35.78
	2018-01-15	35.6	35.75	35.06	35.2
D	2018-01-02	63.29	65.23	63.29	63.78
	2018-01-03	63.78	64.48	63.01	63.48
	2018-01-04	63.32	64.98	63.32	64.54
	2018-01-05	64.79	64.89	61.11	61.59
	2018-01-08	61.4	61.4	58	60.33
	2018-01-09	60	63.4	59.63	63
	2018-01-10	62.61	63.3	62.23	63.01
	2018-01-11	63.01	64.3	62.78	63
	2018-01-12	63.71	63.75	61.5	62
	2018-01-15	61.79	62	59.09	59.4

6 Discussion

The development of economy has promoted the development of stock market, which has gradually become a vital part of the financial market. In turn, it has promoted the further development of economy and become an important driving force of economic development. Stock market has attracted more and more investors' attention. Faced with the rapid development of the stock market, the demand for stock forecasting is becoming stronger and stronger. For investors, stock forecasting is indispensable if they want to seize opportunities in the rapidly changing stock market, effectively avoid risks, and obtain greater benefits. However, the stock market fluctuates sharply and it is difficult to predict [13]. It is found that the future trend of stock can be predicted more accurately through data mining of historical data, so as to provide investors with scientific and reasonable stock selection and help investors to better manage their investments [14].

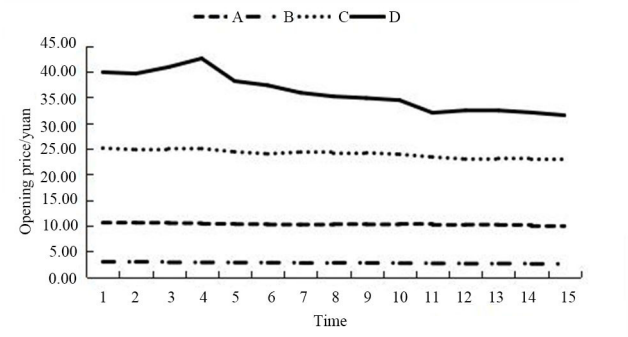

Fig. 8. Prediction of opening price

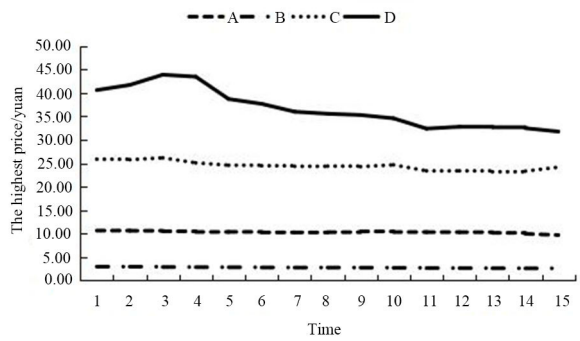

Fig. 9. Prediction of the highest price

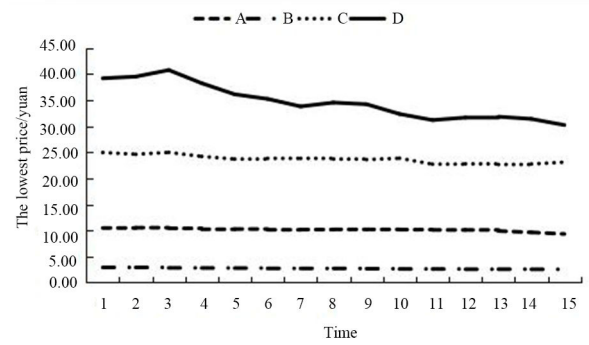

Fig. 10. Prediction of the lowest price

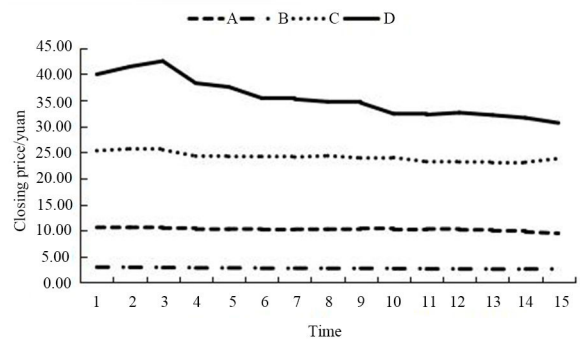

Fig. 11. Prediction of closing price

In this paper, the BPNN method in data mining was studied. Aiming at the slow speed and low precision of BPNN, it was improved by GA. Then the improved prediction algorithm was verified by experiments. A stock was randomly selected from the database and its data were collected. The algorithm was trained using the training samples, and then the opening price, highest price, lowest price and closing price of four stocks were predicted using the trained algorithm. It was found from Fig. 4 to Fig. 7 that the error of the BPNN algorithm was large, there was a large gap between the forecasting result and the actual value, and the data fluctuation was also large. The forecasting error of the opening price using the BPNN algorithm reached 10.59%. The forecasting result curve of the improved BPNN algorithm and the actual result curve fitted well; the prediction errors were small and the average error was 0.8%. It showed that the improved algorithm had higher accuracy than the BPNN algorithm, and it could predict the changes of stock prices more accurately, so as to provide a reliable basis for investors. Then in practical application, the future change law of four stocks were predicted based on the collected data to provide a basis for investors. Moreover the accurate prediction of price stock is beneficial to the stability of the financial market, which can promote the good development of economy.

7 Conclusion

BPNN was improved by GA in this study, and then it was applied in stock forecasting. It was found that the forecasting results were basically consistent with the actual values; the average error of the improved BPNN was about 0.8%, while the average error of BPNN and PSO-BPNN was 7.19% and 4.17% respectively, indicating that the improved BPNN had good forecasting accuracy. In order to provide a basis for investors in stock selection, the future development of four stocks was predicted, and good results were obtained. This work provides some theoretical support for the further promotion and application of the improved BPNN in stock prediction. But there are many factors that can affect the fluctuation of stock price due to the complexity of stock market and the influence of emergency on stock price was not considered. Therefore more factors need to be further considered and the performance of

neural network needs to be further improved to realize more effective stock prediction, which is the direction of future works.

Acknowledgements

This study was supported by Supported by Foundation of Anhui Educational Committee (No: 2015mooc081).

References

[1] B. Majhi, C.M. Anish, Multiobjective optimization based on adaptive models with fuzzy decision making for stock market forecasting, Neurocomputing 167(2015) 502-511.

[2] M. Hu, Y. Sun, A visual stock market trend forecasting method based on state evolution, Journal of Computer-Aided Design and Computer Graphics 26(2)(2014) 302-313.

[3] L. Zhang, Q.Y. Xu, C. Chen, Z.J. Bao, Forecasting stock price based on back propagation neural network, Advanced Materials Research 006-1007(2014) 1031-1034.

[4] M.S.B. Phridviraj, C.V. Gururao, Data mining - past, present and future - a typical survey on data streams, Procedia Technology 12(1)(2014) 255-263.

[5] J. Huang, Y. Zhang, J. Zhang, X. Zhang, A tensor-based sub-mode coordinate algorithm for stock prediction, in: Proc. IEEE Third International Conference on Data Science in Cyberspace, 2018.

[6] H.J. Sadaei, R. Enayatifar, M.H. Lee, M. Mahmud, A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting, Applied Soft Computing 40(C)(2016) 132-149.

[7] J. Wang, R. Hou, C. Wang, L. Shen, Improved - support vector regression model based on variable selection and brain storm optimization for stock price forecasting, Applied Soft Computing 49(2016) 164-178.

[8] B.S. Bini, T. Mathew, Clustering and regression techniques for stock prediction, Procedia Technology 24(2016) 1248-1255.

[9] A. Nayak, M.M. Pai, R.M. Pai, Prediction models for Indian stock market, Procedia Computer Science 89(2016) 441-449.

[10] X. Zhong, D. Enke, A comprehensive cluster and classification mining procedure for daily stock market return forecasting, Neurocomputing (2017) S0925231217310652.

[11] A. Gupta, D.K. Chaudhary, T. Choudhury, Stock prediction using functional link artificial neural network (FLANN), in: Proc. International Conference on Computational Intelligence and Networks, 2018.

[12] A. Izzah, Y.A. Sari, R. Widyastuti, T.A. Cinderatama, Mobile app for stock prediction using Improved Multiple Linear Regression, in: Proc. International Conference on Sustainable Information Engineering and Technology, 2018.

[13] C.Y. Huang, P.K.P. Lin, Application of integrated data mining techniques in stock market forecasting, Cogent Economics & Finance 2(1)(2014) 929505.

[14] A. Sharma, D. Bhuriya, U. Singh, Survey of stock market prediction using machine learning approach, in: Proc. Electronics, Communication and Aerospace Technology, 2017.