Eng Int Syst (2020) 3: 179–183

Engineering
Intelligent Systems

Stock Price Prediction Based on a Neural Network Model and Data Mining

Zheng Fang[1]* and Chaoshin Chiao[2]

[1] *School of Economics & Management, Nanjing University of Science & Technology, Nanjing, Jiangsu 210094, China*
[2] *Department of Finance, National Dong Hwa University, Hualien, Taiwan Corresponding address: No. 200, Xiaolingwei Street, Nanjing, Jiangsu 210094, China*

Rapid economic development has stimulated the development of the stock market, and the existence of the stock market has promoted the flow of the market economy. However, the stock market is risky. An effective and accurate stock price prediction tool can significantly reduce the risk of investors and enterprises. This paper briefly introduces the relevant financial indicators of listed companies that can affect stock prices and a support vector machine (SVM) and Back-Propagation (BP) neural network used for predicting stock prices; the trend of the stock price was then predicted using the SVM combined with the BP neural network. The simulation analysis was carried out on the stock price of an A-share listed company using the MATLAB software. The results showed that the stock price prediction model based on SVM and BP needed less training time than the stock price prediction model based solely on BP. Both models could predict the general trend of the stock price, but the SVM and BP-based prediction model were a better fit for the actual values; the mean square, average absolute percentage error, minimum relative error and maximum relative error also reflected that the combination prediction model was more accurate.

Keywords: Back-Propagation neural network, support vector machine, stock price prediction, financial indicators

1. INTRODUCTION

The rapid development of a market economy has led to the development of the stock market (Maragoudakis & Serpanos, 2016). However, the operation of the stock market is complex and changeable, and it can be affected by objective factors such as market economic policy, the market of listed companies, sudden situations such as a financial crisis and other subjective factors such as the expected psychological changes of shareholders and investors on the stock market (Klibanov & Kuzhuget, 2015). Therefore, the stock market has a considerable risk, and it is almost impossible to make a steady profit in it. The prediction and analysis of the stock price in the stock market is a way to reduce the risk. There are two main methods of stock price analysis, one is fundamental analysis and the other is technical analysis. The former is to determine the current value and future appreciation potential of the stock through the analysis of the business level and development potential of the company to which the stock belongs. The latter is to forecast and analyze the price trend of the existing stock market and the data of stock trading. The two methods have their own merits. However, the calculation of both methods is complex and it is difficult to discover the hidden rules in the data, this ultimately leads to poor prediction and analysis results. The emergence of computer-aided analysis has brought data mining technology, such as association rule analysis, support vector machine and neural networks. With the support of huge computing power, data mining technology can effectively discover the hidden reasons for stock price changes in the stock market and make effective predictions. Chou et al. (2018) proposed a smart time series prediction system based on a

*Email: lfzkaa@163.com

sliding window heuristic optimization to forecast the stock price of the Taiwan Construction Company. The simulation results showed that the prediction system developed had an excellent prediction performance and improved the efficiency of investments. Klibanov et al. (2016) proposed a new empirical mathematical model of the Blake-Scholes equation in order to predict stock prices. In order to verify the validity of the model, 368 randomly selected stock market prices were used. The experimental results showed that the method could effectively select profitable stocks. Nayak et al. (2015) proposed an artificial chemical reaction neural network (ACRN) which predicted the stock market index by training a multi-layer perception model with the artificial chemical reaction optimization. The simulation results showed that the prediction accuracy of ACRN was much higher than that of the multi-layer perception model.

2. SVM

SVM (Das & Padhy, 2018) is a self-supervised learning algorithm in data mining technology. In the training process, the training samples data are those that have been correctly classified by binary classification. Through self-learning of training samples, the SVM model is established. The basic principle of SVM is as follows. The input data is converted into points in n-dimensional space. The values of data are expressed by these coordinates of points in n-dimensional space. Then a hyperplane is found in n-dimensional space to classify the points into two categories. The hyperplane with the best classification effect can be used in the construction of the SVM model. Its construction formula (Zhang et al., 2016) is:

$$\begin{cases} \min \frac{\|W\|^2}{2} \\ y_i \left(W \bullet X_i + b \right) \geq 1 \quad \forall X_i , \end{cases} \tag{1}$$

where W stands for a vector perpendicular to the hyperplane, $y_i(\bullet)$ stands for the distance of the i-th element with the hyperplane in n-dimensional space, X_i stands for the vector of the i-th element in n-dimensional space, and b stands for the constant term of a hyperplane. It can be seen from (1) that the minimum value in the global space can be obtained when the input data of SVM are linearly separable. The objective function of the ideal SVM is shown in equation (1). The hyperplane obtained after calculation can perfectly divide the data vector points in space into two totally inconsistent categories. However, in practical applications, there may be varying degrees of correlation between the data that are input into SVM, and there are collinearities among the vectors. As a result, even if the best hyperplane is obtained by solving the problem and the space is classified, some data vector points will deviate from the correct classification area and be regarded as classification errors. In order to solve the above problems and improve the classification accuracy of the SVM model as much as possible, relaxation variables are introduced on the basis of the original formula to obtain hyperplanes with maximum boundaries and minimum errors:

$$\begin{cases} \min \frac{\|W\|^2}{2} + C \sum_i \varepsilon_i \\ y_i \left(W \bullet X_i + b \right) \geq 1 \quad \forall X_i , \end{cases} \tag{2}$$

where ε_i stands for the slack variable of the i-th element and C stands for the weight of making the distance between the element that is away from the correct classification area and plane minimum when searching for the optimal hyperplane.

The above-mentioned solving process works extremely well for linear separable data, but in reality, the data that are analyzed and predicted by SVM are often non-linear separable. Therefore, before analyzing and predicting actual data with SVM, it is necessary to transform the non-linear separable data into linear separable data. The conversion method is as follows. Firstly, the non-linear separable data are projected into another space, then the data are convolved according to the kernel function to obtain the inner product, and a new linear separable vector is obtained. The commonly used kernel function (Patel et al., 2015) is:

$$K\left(X_i, X_j\right) = \begin{cases} X_i \bullet X_j \\ \left(\alpha X_i \bullet X_j + B\right)^d \\ \tanh\left(\alpha X_i \bullet X_j + B\right) , \end{cases} \tag{3}$$

where X_i, X_j are vectors, α is coefficient, and B is constant.

3. BP NEURAL NETWORK MODEL

The basic structure of a BP neural network (Laboissiere et al., 2015) is divided into an input layer, a hidden layer and an output layer. The hidden layer can be single layer or multi-layer. The BP neural network model is a three-layer structure. x_1, $x_2, \cdots, x_n$ are input vectors X, i.e., the financial indicators of listed companies mentioned above, and D is the output vector, i.e., stock price. In simple terms, the principle of BP neural network is to adjust the weight of the activation function in the hidden layer repeatedly according to the error between the calculated data and the pre-set data, so that the function reflecting the relation between the input data and the hidden layer can be as close to the hidden rule as possible. This is the learning process of BP model. The three most important formulas in the process of learning are: (1) the formula for the forward calculation of actual data; (2) the formula for calculating errors between the actual data and the expected data; (3) the formula for adjusting the weights in the forward calculation formula:

$$d = f\left(\sum_{i=1}^{n} \omega_i X_i - b\right), \tag{4}$$

where d is the actual data calculated by the hidden layer, ω_i stands for the weight of the i-th input index, X_i stands for the i-th input index, n stands for the number of input indicators, b is an adjustment item, and $f(\bullet)$ stands for an activation function (Nayak et al., 2015) used for revealing the rule of the hidden layer.

$$E = \frac{\sum_{k=1} (d_k - y)^2}{2}, \tag{5}$$

where E is the error between the data obtained by the forward calculation and the expected data (Wang et al., 2017), d_k stands for the data obtained from the k-th forward calculation, and y stands for the expected data.

$$\omega_{t+1} = \omega_t + \eta(1 - y)a, \tag{6}$$

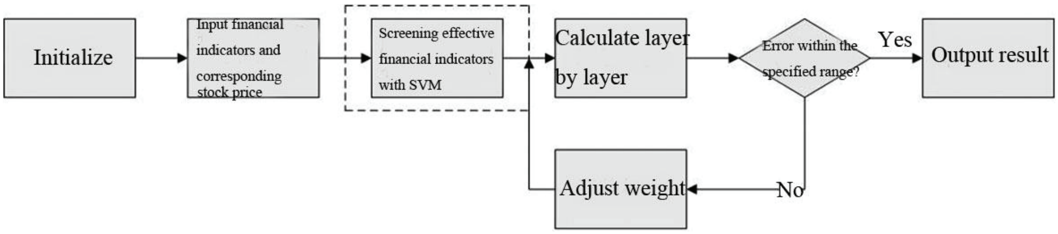

Figure 1 The learning process of BP neural network combined with SVM

Table 1 Relevant indicators affecting stock price trends.

Number	Index	Unit
X_1	Operating profit ratio	%
X_2	Rate of return on total assets	%
X_3	Cash flow ratio	%
X_4	Debt coverage rate	%
X_5	Payable receivable turnover rate	%
X_6	Asset turnover rate	%
X_7	Equity turnover rate	%
X_8	Asset-liability ratio	%
X_9	Sustainable growth rate of assets	%
X_{10}	Capital maintenance rate	%

where ω_{t+1} refers to the weight after one step of learning (Kristjanpoller & Minutolo, 2015), ω_t stands for the weight before learning, and η stands for the learning step length, also called the learning rate.

The learning process of a BP neural network combined with SVM is shown in Figure 1.

1. Initialize the weight to the minimum value.
2. Input the financial index data and the corresponding stock price data.
3. SVM is used to screen the company's financial indicators, the indicators that have the greatest impact on the stock price trend are selected; calculation is then performed layer by layer according to the forward propagation direction of the network.
4. After calculating the stock price layer by layer, it is compared with the actual stock price. If the formula is correct, the result will be within the prescribed range. If the result is not within the prescribed range, the weight of the calculation formula in the hidden layer and the output layer will be adjusted. After the weight adjustment, the error of the actual output vector and the expected output vector will be recalculated, and the above steps will be repeated until the result reaches the prescribed range.

4. EXAMPLE ANALYSIS

4.1 Experimental Environment

MATLAB software (Dash & Dash, 2016) was used to compile a BP neural network model algorithm and an SVM algorithm. The experiment was carried out on the laboratory server which has a Windows 7 operating system, an Intel i7 processor and 16GB of memory.

4.2 Experimental Data

An A-share listed company was selected from the CSMAR database according to the following conditions: (1) non-ST stock; (2) complete financial data of the listed company for the past ten years; (3) no suspension or delisting within the period of the prediction analysis in this study. After the selection of the listed company, 10 of the financial indicators mentioned above and corresponding stock prices were selected from the company's quarterly, semi-annual and annual financial reports in 2010–2017 as training samples; 10 financial indicators, as shown in Table 1, were then selected from the financial report of the first quarter of 2008 as testing samples, and the corresponding stock prices were used as real values to measure the accuracy of the model.

4.3 Experimental Results

As shown in Figure 2, the mean square error of the stock price prediction model based on the BP neural network algorithm and the stock price prediction model based on the SVM and BP neural network algorithm both showed a decreasing trend with the increase of training times until finally the mean square error stabilized at 10^{-5}. Although both models improved in accuracy with the increase of training, the stock price prediction model based on the BP neural network needed over 800 training runs to stabilize the mean square error at 10^{-5}, while the stock price prediction model based on the SVM and BP neural network only needed 600 training runs to reach the same value. Moreover, the curve changes in

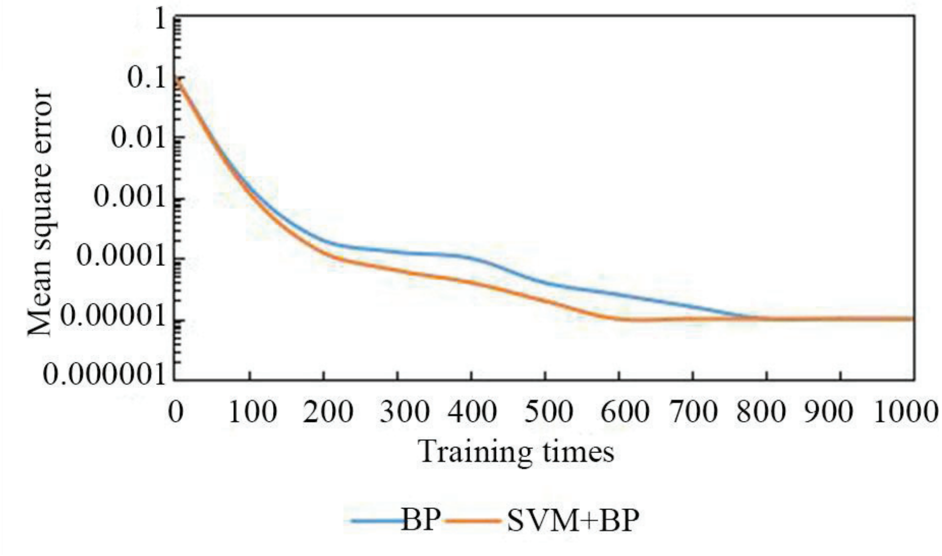

Figure 2 Changes of mean square error during training of two models

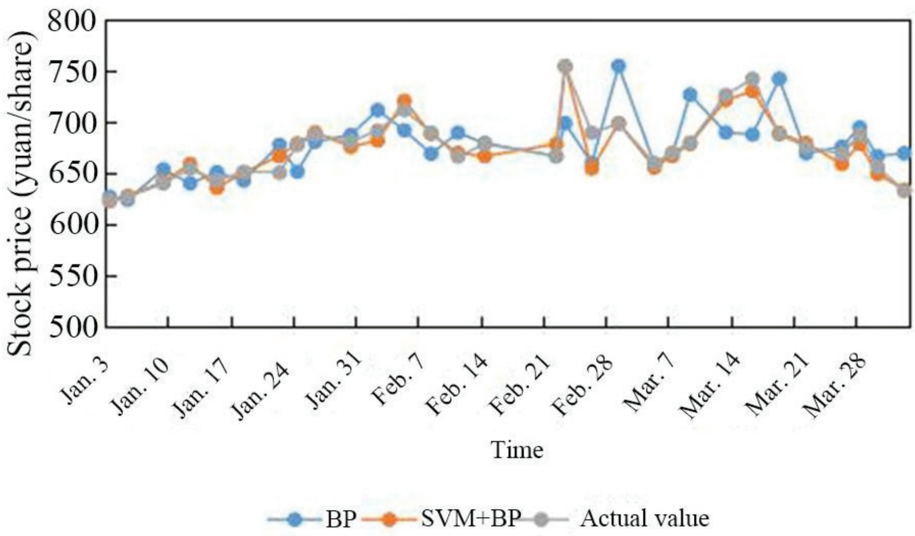

Figure 3 Comparison of the predicted and real values of the two models

Table 2 Accuracy comparison of two models.

Model	Mean square error	Mean absolute percent error	Minimum relative error	Maximum relative error
BP	66.125	1.1243	0.0028	0.0218
SVM+BP	55.871	1.0003	0.0014	0.0211

Figure 2 showed that the stock price prediction model based on the SVM and BP neural network algorithm had a faster training convergence speed than the BP-based stock price prediction model. The reason for this is that the input financial indicators were classified and screened by SVM in the stock price forecasting model based on SVM and BP and the key indicators affecting the stock price were selected, this greatly reduced the number of input vectors and made the convergence speed of the model faster in training.

As shown in Figure 3, the stock price of the listed company in the first quarter of 2018 was relatively stable, with significant increases in mid-February and early March. Overall, the predicted values of the two models were close to the real value, and the trend was roughly the same. The predicted value of the stock price prediction model based on the SVM and BP neural network was closer to the real value than the stock price prediction model based on the BP neural network.

Table 2 clearly shows that the mean square error, mean absolute percent error, minimum relative error and maximum relative error of the prediction model based on the SVM and BP neural network model were all smaller than those of the model based solely on BP. From the comparison, it was found that the prediction model based on the SVM and BP neural network model had a significantly better prediction effect. The reason for this is that the SVM algorithm eliminated the financial indicators that had little impact on stock prices, which reduced the fluctuation of the predicted value and improved prediction accuracy.

5. CONCLUSION

This paper briefly introduced the relevant financial indicators of listed companies that can affect stock prices and the SVM and BP neural network algorithm for predicting a trend in stock price. The combination of the SVM and BP neural network algorithm was then used to predict a stock price trend. The stock price of an A-share listed company was simulated and analyzed using MATLAB software. After 800 training runs, the BP-based stock price prediction model was stable at 10^{-5}, and the stock price prediction model based on the SVM and BP neural network was stable at 10^{-5} after only 600 training runs. Both prediction models could predict the stock price trend, and the prediction performance of the model based on the SVM and BP neural network was more accurate.

REFERENCES

1. Chou, J.S. & Nguyen, T.K. (2018). Forward Forecast of Stock Price Using Sliding-Window Metaheuristic-Optimized Machine Learning Regression. *IEEE Transactions on Industrial Informatics*, 14(7), 3132–3142.
2. Das, S.P. & Padhy, S. (2018). A Novel Hybrid Model Using Teaching-Learning-Based Optimization and a Support Vector Machine for Commodity Futures Index Forecasting. *International Journal of Machine Learning & Cybernetics*, 9(1), 97–111.
3. Dash, R. & Dash, P.K. (2016). Efficient Stock Price Prediction Using a Self Evolving Recurrent Neuro-Fuzzy Inference System Optimized Through a Modified Differential Harmony Search Technology. *Expert Systems with Applications*, 52, 75–90.
4. Huang, C.H., Yo, J.L. & Chen, A.P. (2015). Using the Event Classifier System to Forecast the Stock Price: An Empirical Study of Taiwan Financial Market. *Applied Mechanics and Materials*, 764–765, 7.
5. Klibanov, M.V. & Kuzhuget, A.V. (2015). Profitable Forecast of Prices of Stock Options on Real Market Data via the Solution of an Ill-Posed Problem for the Black-Scholes Equation. *Papers*, 32(1).
6. Klibanov, M.V., Kuzhuget, A.V. & Golubnichiy, K.V. (2016). An Ill-Posed Problem for the Black-Scholes Equation for a Profitable Forecast of Prices of Stock Options on Real Market Data. *Inverse Problems*, 32(1): 015010.
7. Kristjanpoller, W. & Minutolo, M.C. (2015). Gold Price Volatility: A Forecasting Approach Using the Artificial Neural Network GARCH Model. *Expert Systems with Applications*, 42(20), 7245–7251.
8. Laboissiere, L.A., Fernandes, R.A.S. & Lage, G. (2015). Maximum and Minimum Stock Price Forecasting of Brazilian Power Distribution Companies Based on Artificial Neural Networks. *Applied Soft Computing*, 35, 66–74.
9. Maragoudakis, M. & Serpanos, D. (2016). Exploiting Financial News and Social Media Opinions for Stock Market Analysis using MCMC Bayesian Inference. *Computational Economics*, 47(4), 589–622.
10. Nayak, S.C., Misra, B. & Behera, H.S. (2015). Artificial Chemical Reaction Optimization of Neural Networks for Efficient Prediction of Stock Market Indices. *Ain Shams Engineering Journal*, 8(3).
11. Patel, J., Shah, S., Thakkar, P., et al. (2015). Predicting Stock and Stock Price Index Movement Using Trend Deterministic Data Preparation and Machine Learning Techniques. *Expert Systems with Applications An International Journal*, 42(1), 259–268.
12. Thenmozhi, M. & Chand, G.S. (2016). Forecasting Stock Returns Based on Information Transmission Across Global Markets Using Support Vector Machines. *Neural Computing & Applications*, 27(4), 805–824.
13. Wang, D., Guo, H., Luo, H., et al. (2017). Multi-Step-Ahead Electricity Price Forecasting Using a Hybrid Model Based on Two-Layer Decomposition Technology and BP Neural Network Optimization by Firefly Algorithm. *Applied Energy*, 190, 390–407.
14. Zhang, X.D., Li, A. & Ran, P. (2016). Stock Trend Prediction Based on a New Status Box Method and Adaboost Probabilistic Support Vector Machine. *Applied Soft Computing*, 49, 385–398.