The share price trend prediction of China's coal listed companies —Based on PCA & BP neural network

Lanjin Yang

School of Humanities and Economic Management, China University of Geosciences (Beijing), Beijing, 100083, China

yanglanjin@163.com

Abstract. This paper starts from the financial indexes of the coal listed companies and use financial indexes which represent cash flow capacity, solvency, operational capability, profitability and development ability to predict share price by BP neural network. Firstly, the input variables are optimized by principal component analysis, and then the trained network is saved. Finally, the stock price of coal is forecasted, and the results are analyzed.

Keywords: coal; listed company; stock price; principal component analysis; BP neural network.

1. Introduction

Coal accounted for about 2/3 of the China's primary energy structure and it plays an important role in the national economy. The stock price of coal listed company is a barometer of the development of the coal industry so it is important to research how to predict the stock price of coal listed companies [1]. It not only enables us to understand the development trend of the coal industry in advance, but also provides an important reference for investors in the coal stocks. The basic factors affecting the stock price mainly include the company's operating conditions, industry and sector factors and macroeconomic factors. In all factors affecting the stock price, the most can be directly understood by investors is the financial report which reflects the operating conditions of listed companies. And in the case of the same systemic risk, the core factor that affects the share price of the same industry listed companies is the financial situation. This paper tries to use the method of PCA&BP (principal component&back propagation) neural network to forecast the stock price of coal listed companies based on the financial index of coal listed companies.

2. Introduction of the Method

2.1 Principal Component Analysis

Principal component analysis, namely PCA, is a kind of exploratory statistical analysis method, which concentrates information distributed over a set of variables on a few comprehensive indexes (principal components).

When we predict the stock price, we choose too many indicators which influence the stock price. The BP neural network has lots of disadvantages such as the too many input layers, the complex structure , the too long convergence and the low efficiency of operation. The combination of PCA and BP neural network can simplify these problems. The steps of principal component analysis are as follows. (1) Index data standardization. (2) Calculate covariance matrix of normalized sample correlation matrix. (3) Calculate the eigenvalues of the covariance matrix and the corresponding characteristic vectors. (4) Determine the number of principal components. When determining the number of principal components, we can adhere to the principle of characteristic value greater than 1, or adhere to the principle of cumulative contribution rate of more than 85%.

2.2 BP Neural Network

The BP neural network is a kind of multilayer feedforward network trained by the error back propagation algorithm, which is one of the most widely used neural network models. Its learning rule is to use the steepest descent method. And it minimizes error sum of squares of the network by reverse spread which constantly adjust the network weights and thresholds. The typical BP neural network

model topology consists of an input layer, one or more hidden layers and an output layer. The learning process of BP neural network is illustrated with a simple supervised learning process, as shown in figure 1. BP neural network algorithm is divided into two steps, namely, the forward and reverse propagation.
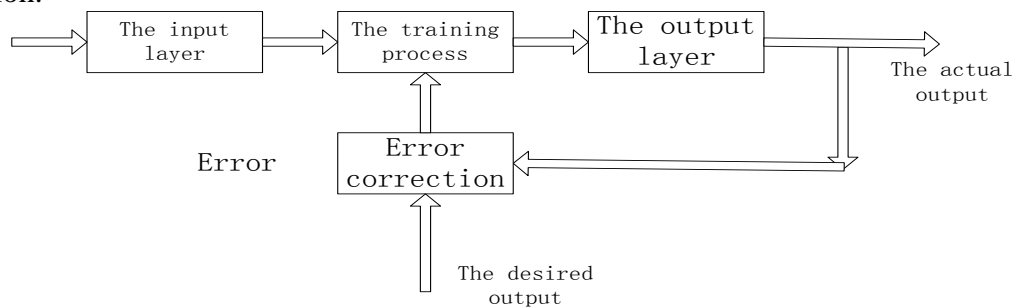

Fig. 1 Learning block diagram of perceptron

3. Index Selection and Research Process

3.1 Index Selection

The financial reports of the listed company should be as comprehensive as possible to reflect the operating conditions of the enterprise, so there are too many indicators in the financial reports. In order to make the results more accurate and persuasive, this paper follows the following principles when selecting representative indicators from a large number of financial indicators: (1) comprehensive principle. Comprehensive financial indicators can be more accurately reflect the company's operating status of listed coal companies. Therefore, when we select indicators, we should take into account five aspects of financial indicators which reflect cash flow capacity, solvency, operational capability, profitability and development ability. We should select the most representative indicators of these five aspects, so that the results of the study are more accurate. (2) Operational principle. Operational principle means the indicators are easy to operate. First, we should ensure that the index data source is true and reliable and then we should ensure the availability of data. We cannot select the financial indicators which cannot be obtained even if they may well reflect the financial situation of listed companies. Growth rate of net profit can be more adequately reflect the development ability of coal listed companies, but when net profit for the last year is negative, then this year's growth rate of net profit cannot be calculated. Based on operational principle, this index cannot be selected to reflect the development capacity of coal listed companies

Based on the above two principles and referred to other people's literature [2-4] , this paper select operating cash flow per share and cash flow ratio to reflect cash flow capacity; flow ratio to reflect solvency; inventory turnover ratio, accounts receivable turnover and fixed assets turnover ratio to reflect operational capability; earnings per share, main business profit margin and rate of return on net assets to reflect profitability; growth rate of total assets and growth rate of main business revenue to reflect development ability.

3.2 Research Process

The research object of this paper is the stock price of coal listed companies. The sample companies belong to the same industry, so we ignore the impact of macroeconomic and industry factors on stock prices. We only consider the operating conditions of the listed companies on the stock price. First, this paper collect operating cash flow per share, cash flow ratio, flow ratio, inventory turnover ratio, accounts receivable turnover, fixed assets turnover ratio, earnings per share, main business profit margin, rate of return on net assets, growth rate of total assets, growth rate of main business revenue these eleven indicators and then use the SPSS21.0 software to carry out the principal component analysis of these 11 Financial indicators. Finally, this paper make the main components of the 11 financial indicators as input variables and the corresponding stock price as the output variable When using the method of BP neural network to predict the stock price.

3.3 Samples and Data Sources

In this paper, the sample excludes ST shares of listed coal companies and the listed companies which cannot get selected financial indicators. This paper select 30 coal listed companies and the financial indicators data from Sina Finance. This paper obtains the closing price of samples I on April 29, 2016 according to flush software stock k line graph. The annual reports of listed companies in 2015 should be publish before April 30, 2016, so the closing share price on April 30, 2016 is more able to reflect the financial situation of listed companies in 2015.Since April 30, 2016 is not a trading day, so we select closing price of the stock on April 29th.

4. Empirical Analysis

4.1 Results of Principal Component Analysis

This paper establish a M30-11 matrix based on the 11 financial indicators of 30 coal listed companies, and analyze the main components of the original datas by using SPSS21.0 software. From table 1 we can know that the first four principal components of the characteristic value which is greater than 1 accounts for 81.045% of the cumulative contribution rate. So we can think that the first four principal components can represent the vast majority of information on the company's operating conditions.

Table 1 the analysis table of principal component extraction

Component	Initial Eigenvalues		
	Eigenvalues	%of Variance	Cumulative%
1	3.201	29.098	29.098
2	2.391	21.737	50.835
3	1.983	18.028	68.863
4	1.340	12.182	81.045
5	0.804	7.313	88.358
6	0.470	4.274	92.632
7	0.320	2.909	95.541
8	0.239	2.172	97.712
9	0.140	1.273	98.986
10	0.102	0.931	99.916
11	0.009	0.084	100.000

From the Table 2, we can know that in the first principal component, earnings per share, main business profit margin and rate of return on net assets yield of the load is much greater than other financial indicators of the load. So the first principal component is mainly explained by earnings per share, main business profit margin and rate of return on net assets, which represents the company's profitability. Similarly, the second principal component is mainly explained by operating cash flow per share, cash flow ratio, flow ratio and fixed assets turnover ratio which represent the company's cash flow capacity, solvency and operational capability of fixed assets. The third principal component is mainly explained by growth rate of total assets and growth rate of main business revenue, which represents the company's development ability. The last principal component is mainly explained by inventory turnover ratio and accounts receivable turnover, which represents the company's operational capability of the current assets. These Four principal components are input variables of BP neural network.

When the four principal components extracted by principal component analysis, we can explain all the variables with fewer variables. And at the same time the high correlation between variables is converted into independent or uncorrelated variables. The managers and investors of the coal listed companies can assess the company's stock price reasonable by detecting these indicators.

Table 2 Component Matrixa

Indicator	Principal Component			
	1	2	3	4
Operating Cash Flow Per Share	0.505	-0.670	0.083	0.037
Cash Flow Ratio	0.381	-0.793	0.143	-0.098
Flow Ratio	0.366	0.732	-0.129	-0.062
Inventory Turnover Ratio	-0.419	-0.031	0.002	0.744
Accounts Receivable Turnover	-0.103	-0.308	-0.059	0.793
Fixed Assets Turnover Ratio	0.252	0.795	0.000	0.282
Earnings Per Share	0.911	0.068	-0.145	0.139
Main Business Profit Margin	0.833	-0.063	-0.153	0.058
Rate of Return on Net Assets	0.915	0.081	-0.118	0.186
Growth Rate of Total Assets	0.116	0.117	0.981	0.029
Growth Rate of Main Business Revenue	0.207	0.147	0.957	0.074

4.2 The Establishment of BP Neural Network

This paper selects a three-layer-BP neural network to predict the stock price. The neural network learning generally requires repeated trainings so that the error value gradually to zero. In this case the output will be consistent with the desired. So neural network learning need consume a certain amount of time. At present, it is a very important research topic to improve the learning speed of neural network and to reduce the number of repeated learning, which is also the key problem in real time control [5, 6].

4.3 Forecast result

In this paper, the neural network input layer contains four neurons, that is, the principal component analysis of the four main components of the score; the hidden layer of seven neurons; the output layer of a neuron, that is, the prediction of the stock price. The first 25 samples were used as training samples, and then 5 samples were used as test samples. Referring to other people's papers [7], the BP neural network adopts 4×8×1 structure, in which the error position is 0.005, the learning rate is 0.3, the maximum value of training times is 1000. The predicted results and the actual stock price are basically the same by the BP neural network. And the forecast results of the five companies share price as shown in figure 2. The average correct rate for the shares of the five companies reached 76.32%.
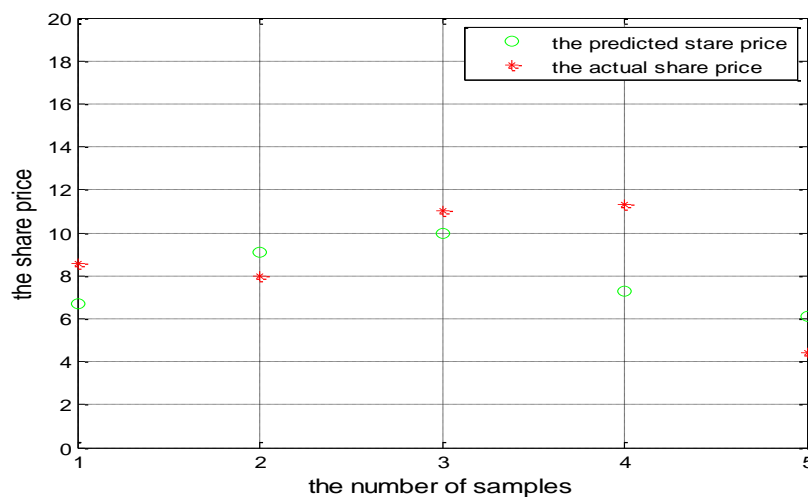

Fig. 2 The share price

The predicted results show that the PCA & BP neural network method can effectively realize the prediction of the stock price in coal industry.

5. Conclusion

This paper starts from all levels of coal listed company's financial indicators and predicts share price of coal listed companies. According to the empirical result, the combination of PCA and BP neural network method can effectively reduce the input layers and make the result more accurate. Through the method of PCA&BP neural network, the forecast accuracy of the stock price of coal listed companies is higher. If this method is applied to the prediction of the stock price of listed coal companies, the investors of coal listed companies stock can understand that the company's stock price is undervalued. At the same time, we can understand the future development trend of the coal industry.

References

[1] Peng Yu, Wang Dong. Study on Investment Analysis on Stock Price of Listed Companies in coal Industry in China [J]. Coal Economic Research. Vol. 32 (2012) No. 5, p.51-53.

[2] Huang Mei, Kang Dexiang, Deng Chengchao. The application of accounting information in the forecast of stock price movement [J]. Securities. (2008) No. 1, p. 169-171

[3] Zhu Yongming, Shao gengyun. The trend of stock price forecasting based on BP neural network [J]. Finance and Accounting Monthly. (2013) No. 7, p. 76-78

[4] Liu Zhiyong. Correlation analysis of financial index and stock price fluctuation of listed companies [J]. China Price. (2015) No. 11, p. 54-56.

[5] Chang Huang. Research on Underwater Target Recognition Based on BP Neural Network [D]. Shen yang: Northeastern University, 2010.

[6] Chen Si. Learning Rate Parameter Improve Methods for BP Neutral Network [J]. Journal of Changchun Normal University (Natural Science). Vol. 29 (2010) No. 1, p. 26-28.

[7] Zeng Xiangxiao, Zhao Liangzhong. Prediction model of quality of wine on principal component analysis and BP neural networks [J]. Food and machinery. Vol. 30 (2014) No. 1, p. 40-44.