

# Research on the Application of Data Mining Technology in Stock Market Development Prediction

Wenjing Gong<sup>1, a</sup>

<sup>1</sup>St. Francis College, 180 Remsen Street, Brooklyn, NY, 11201, USA

<sup>a</sup>wenjinggong@hotmail.com

**Keywords:** Data Mining, Stock Prediction, Clustering, Neural Network

**Abstract.** The stock market is an important part of the securities industry and financial industry essential, universal attention by investors. Effective stock prediction plays an important role in the field of financial investment, so on the stock price of analysis and prediction has a very important theoretical significance and practical value. Based on data mining technology, this paper analyzes and predicts the trend of stock market. The goal is to establish a forecasting model using neural network in data mining, and to seek the combination of data mining algorithm and stock forecast through the analysis of the forecast process and the forecast results.

## Introduction

As an important feature of the market economy in the stock market, about from the birth of tens of millions of investors in the heart. High risk and high returns are the characteristics of the stock market, so investors are concerned about the stock market, the analysis of the stock market, trying to predict the development trend of the stock market. For more than 100 years, some analytical methods with the emergence and development of the stock market to gradually improve, such as Dow analysis method, K line graph analysis method, histogram analysis method, point and figure chart analysis method, moving average method, and morphological analysis, trend analysis method, the angle analysis method, the mysterious series with gold segmentation than spiral calendar, four space method [1-2], along with the computer technology in the stock analysis of the popularization and application of the field, has introduced new index analysis method. However, strictly speaking these methods are just the means of analysis, and it cannot directly predict the dynamics of the stock market [3].

In addition, it is also trying to establish a model to forecast the stock market by means of regression analysis and other statistical methods. It is a very large amount of data to be processed by the traditional forecasting technology. Data mining technology research work has made great progress, all kinds of data mining technology greatly promotes the people to analyze and deal with a large amount of data information ability and people brought good economic benefit. Therefore, we can foresee the data mining technology in the stock market pretest will have great potential.

## The basic technology of data mining

Data mining is a long-term research and applications of database technology is an inevitable result. At the same time, the development of database technology more advanced stage, which can not only a large amount of historical data, the potential links can also be found in the unknown history data query. In the face of a large amount of data in the current state, the association rules is an important branch of data mining, as a kind of advanced and intelligent data processing and analysis technology has become a hot spot. By association rules mining, it can be implied in the sea have the potential value of data useful information. The goal of association rules is to extract the most interesting patterns from the effective method. Data mining is a kind of decision support process, it mainly based on artificial intelligence, machine learning, pattern recognition, statistics, database, visualization technology, high degree of automation, data analysis of enterprise, make inductive reasoning, dig out the potential of the model help users to adjust market strategy, reduce

the risk, to make correct decisions. Data mining process from a technical point of view, as shown in figure 1.

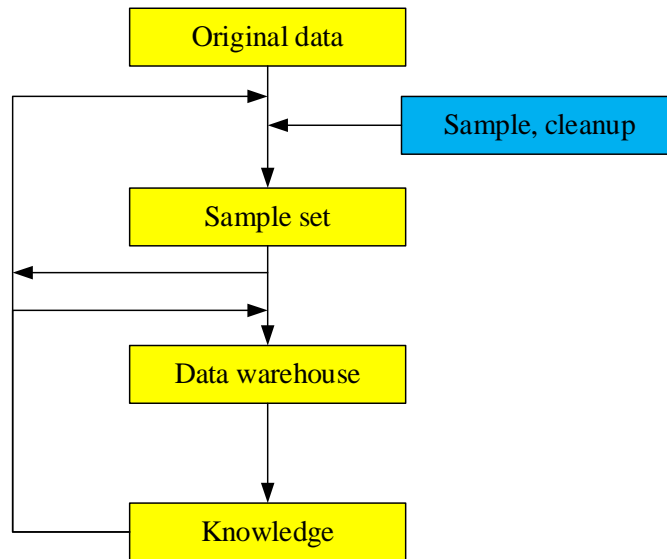


Figure 1. The basic process of data mining

After data collection, it need sampling and cleanup. Cleanup is the result of an example data set. Data warehouse is an effective data storage format, which is very good for data mining. It can use a variety of data mining algorithms, such as neural network algorithm. The final phase is sometimes required to return to the above process.

### The proposed approach of BP neural network

Neural network is to imitate and expansion of human brain cognitive function model of intelligent information processing system, which is composed of a large number of simple processing units (neurons) is a complex dynamic nonlinear adaptive learning system. Artificial neural network as a new type of artificial intelligence technology is one of the research directions in the world. The rapid development of artificial neural networks with a large number of neurons extensively interconnected [4-5]. Constitute a complex network system, in some degree, it simulates the cerebral nervous system of information processing, storage and retrieval functions, learning, memory and computing intelligence capabilities. Artificial neural network optimization, signal processing and pattern recognition, intelligent control, fault diagnosis and many other fields have broad application prospects.

The ability of the BP neural network to reverse propagation and correct the error mapping function. After the input and output parameters of the unknown system, the associative memory system can be expressed. The learning process of BP neural network is composed of two parts, signal propagation and error back propagation. Forward propagation is represented as an input sample input from the input layer to the hidden layer to the output layer. If the output layer does not conform to the desired output, the output error back propagation step by step through the hidden layer to the input layer will affect the error propagation to all the nodes in each layer is connected with the right weight for each node is shown in Figure 2.

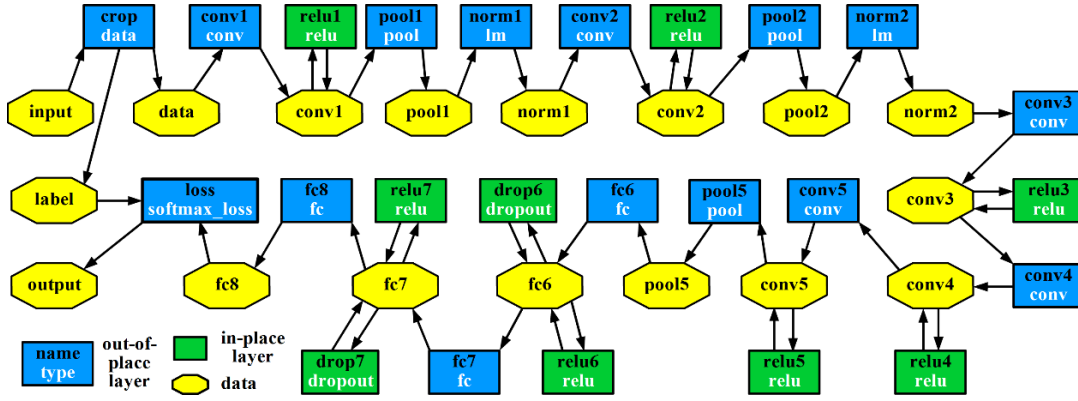


Figure 2. The general structure of the proposed neural network

Note that the input to the filter learning phase is normalized power spectrum, it is adopted the use of the positive index of logarithm. The following formula 1-3 shows this process. Equation 1 shows each nodes in the hidden layer of input. Expression 2 to the output of hidden layer nodes. Formula 3 shows the final output.

$$S_j = \sum_{i=1}^m w_{ij}x_i - \theta_j \quad (j=1,2,...,p) \quad (1)$$

$$b_j = 1 / \left[ 1 + \exp \left( - \sum_{i=1}^m w_{ij}x_i + \theta_j \right) \right] \quad (j=1,2,...,p) \quad (2)$$

$$x_{i+1} = 1 / \left[ 1 + \exp \left( - \sum_{j=1}^p v_j b_j + \gamma \right) \right] \quad (3)$$

In terms of data based methods, the incremental learning neural network model provides many benefits due to the robust-ness in dealing with large data sets and distributed learning capability. BP neural network in training before, each layer of the connection weights and thresholds randomly initialized with values in [0,1], which means this method is not random initialization and optimization tend to make the convergence speed of the BP neural network is slow and prone to make the final results for the optimal solution. Using genetic algorithm to optimize the initial weights and the threshold distribution of the initial weights and thresholds can make the BP neural network has a higher accuracy.

### The application of data mining in NASDAQ index Prediction

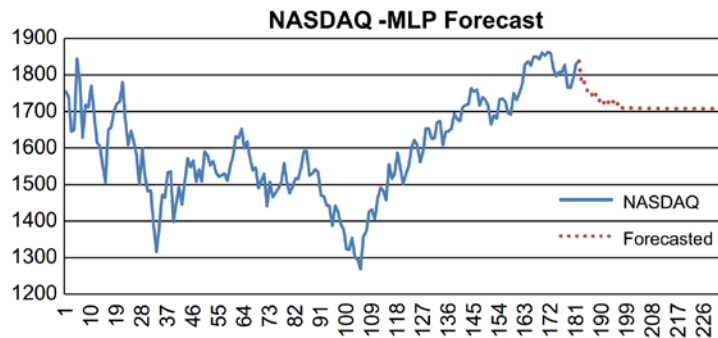


Figure 3. The training data of NASDAQ market and forecast

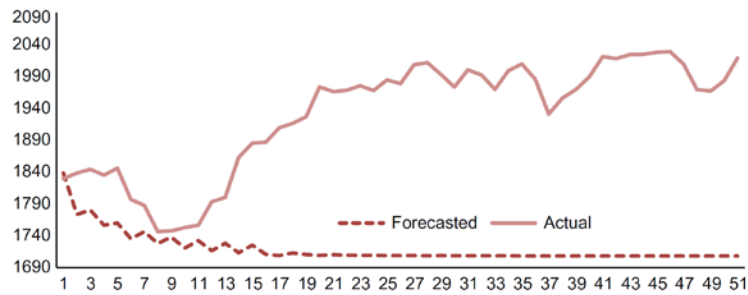


Figure 4. Test results of NASDAQ market based on data mining

The NASDAQ market in the daily exchange of this research is from October 7, 2012 to June 26, 2013. The first 146 days are used for training and cross validation and continuous 36 for testing. In Figure 3, the training data and the forecast period are given, and the test results based on the proposed neural network are shown in Figure 4.

## Conclusion and Summary

Data mining as an interdisciplinary, involving database, statistics, artificial intelligence and machine learning, and other fields, with the development of computer technology, data and information explosion, the data mining technology is more and more use of its powerful information discovery and prediction ability, and provide important information and knowledge. This paper introduces data mining algorithm, especially the basic principle of neural network algorithm for data mining. In the empirical part, this article around the market timing problem, stock index prediction method was proposed based on neural network. With the development of data mining, and its application in the securities industry will be more and more widely, and more and more to the intelligence development, its classification and prediction accuracy will also unceasingly to enhance.

## References

- [1] Liao S H, Chou S Y. Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio[J]. *Expert Systems with Applications*, 2013, 40(5): 1542-1554.
- [2] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market[J]. *Journal of Computational Science*, 2011, 2(1): 1-8.
- [3] Chang P C. A novel model by evolving partially connected neural network for stock price trend forecasting[J]. *Expert Systems with Applications*, 2012, 39(1): 611-620.
- [4] Kannan K S, Sekar P S, Sathik M M, et al. Financial stock market forecast using data mining techniques[C]//*Proceedings of the International Multiconference of Engineers and computer scientists*. 2010, 1.
- [5] Dutta A, Bandopadhyay G, Sengupta S. Prediction of stock performance in the Indian stock market using logistic regression[J]. *International Journal of Business and Information*, 2012, 7(1): 105-136.
- [6] Preethi G, Santhi B. STOCK MARKET FORECASTING TECHNIQUES: A SURVEY[J]. *Journal of Theoretical & Applied Information Technology*, 2012, 46(1).
- [7] Ticknor J L. A Bayesian regularized artificial neural network for stock market forecasting[J]. *Expert Systems with Applications*, 2013, 40(14): 5501-5506.