Data Mining Privacy Preservation Techniques in Stock Market Data Analysis

Nirmal Singh [1]

Stock market forecasting is an important financial subject that has been attracted many researchers and suddenly from last year researchers focuses on it. There is an assumption that the analysis of fundamental information publicly available from the past has some predictive relationships to the future stock returns. Therefore we can predict the stock market variations such that which companies share value in next time will go high or low. This technology is designed to help investors to discover hidden patterns from the historic data that have probable predictive capability in their investment decisions. The prediction of stock markets is a challenging task of financial time series prediction. Data analysis is one way of predicting if future stocks prices will increase or decrease. Five indications analyzing stocks were combined to predict if the day's closing price would increase or decrease. This paper discussed various data mining techniques which are able to predict with future closing stock price will increase or decrease better than level of significance. To predict stock market data we are using Association Rules, Neural Network and Pattern recognition techniques. It supports numerically and graphically.

ISSN 2395-0226

Introduction

The stock market is basically a non-linear, nonparametric system that is extremely difficult to model with any reasonable accuracy [1]. Investors have been trying to find a way to predict stock prices and to find the right stocks and right timing to buy or sell. To achieve those objectives, and according to [2], [3-4] some research used the techniques of fundamental analysis, where trading rules are developed based on the information associated with macroeconomics, industry, and company. The authors of [5] and [6] said that fundamental analysis assumes that the price of a stock depends on its intrinsic value and expected return on investment. According to the market analysis most of the people believes that fundamental analysis is a good method only on a long-term basis. However short- and medium term schedule the fundamental analysis is generally not suitable.

Some other research used the techniques of technical analysis [2], in which trading rules were developed based on the historical data of stock trading price and volume. Technical analysis as illustrated in [5] and [7] refers to the various methods that aim to predict future price movements using past stock prices and volume information. It is based on the assumption that history repeats itself and that future market directions can be determined by examining historical price data. Thus, it is assumed that price trends and patterns exist that can be identified and utilized for profit. Most of the techniques used in technical analysis are highly subjective in nature and have been shown not to be statistically valid.

Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area. Data mining refers to extracting or mining knowledge from large data stores or sets. Some of its functionalities are the discovery of concept or class descriptions, associations and correlations, classification, prediction, clustering, trend analysis, outlier and deviation analysis, and similarity analysis. Data classification can be done in many different methods; one of those methods is the classification by using Decision Tree. It is a graphical representation of all possible outcomes and the paths by which they may be reached.

Data mining is a method of extracting unknown projecting information from large databases which is a widespread technology that helps organizations to focus on the most important information in data repositories with great potential Decision trees and artificial neural networks can be trained by using an appropriate learning algorithm.

Data mining

Data mining tasks can be classified as follows:

Analysis of survey data

It is an interactive and visual method of easily view data without a clear idea of what we are looking for. • Explanatory simulation: It includes a model for the whole probability information dissemination to express all the data; it divides large dimensional space into groups and models that describes the relationship between variables. • Analytical modelling: This model allows to predicting the value of one variable on the identified values of other variables [7].

Identify patterns and rules

It refers to the model of detection, recognition and the goal is to

identifying fraud in the field of space determined by various types of agreements in which data points are significantly different from others. • Acquisition summary: This task is often used for text and image data sets that find the scheme to extract the data that is parallel to the structure of interest in the data set.

Some of the data mining systems are as follows:

Statistical Data Mining

Statistics provide a useful tool for data mining and they can be used to analyze or make inferences about data to discover useful patterns from a dataset. The database is integrated with statistical functions to draw statistical conclusions about the dataset in the database.

Neural network for Data Mining

In traditional DBMS, Data is stored as a shape of structured records. When any query is submitted, database system searches for and retrieves records that match user's query criteria. Artificial neural network offers an excellent way for the recognition of intelligent query processing in large databases, especially for data retrieval and knowledge extraction based on partial matches. Neural network uses different methods. It does not need to identify empirical rules in order to make predictions. A neural network generates a network by examining a database and by identifying and mapping all significant patterns and relationships that exist among different attributes. The network then uses a particular pattern to predict an outcome. The neural network tries to identify an individual mix of attributes that reveals a particular pattern. This process is repeated using a lot of training data, consequently making changes to the weights of the data for more accurate pattern matches. The patterns that exist among the attributes in the database can be identified, and the influence of each attribute can be quantified. Neural network concentrate on identifying these patterns

Three different types of datasets used in this process are as follows:

Training set

It is used for training and for teaching the network to recognize patterns.

Validation set

A set of examples is used to tune the parameters of a classifier by choosing the number of unknown nodes in a neural network.

Test set

Neural Network performance is tested using this. It consists of a set of examples used only to assess the performance of a fully specified classifier.

Clustering

It can be used as a data-mining method to group together items in a database with similar characteristics. It is on how to group data items based on the similarities among them. A cluster is a set of data items grouped together according to common properties and

is considered an entity separate from other clusters. Hence, a database can be viewed as a set of multiple clusters for simplified processing of data analysis. It can be used to achieve objectives such as "identifying critical business values" or "discovering interesting patterns from the database."

Fuzzy sets for data mining

It is a general case of an arbitrary set. It is a set without a crisp boundary. While conventional sets have only two possible values, 0 and 1, fuzzy sets do not have this arbitrary boundary to separate members from non members. It can be used to describe everyday business applications. In the real world, problems are often vague and imprecise, so they cannot be described in the conventional dual (true or false) logic ways. But, Fuzzy logic allows a continuous gradation of truth values ranging from false to true in the description process of application models.

Litrature survey

Marijana Zekic [9] showed that the neural network (NN) accuracy mostly ranges from 70%‑80%. NNs outperform statistical methods for a 5%-20% higher accuracy. The author claimed that if NN is combined with expert system, it will perform higher accuracy rate than only NN.

Bruce J. Vanstone, Gavin Finnie and Clarence Tan [10] demonstrated that artificial neural network (ANN) can be used to identify stocks with a potential to rise significantly on the basis of the stocks fundamental attributes. They have also found that the ANN configurations outperformed their non-neural equivalents. The majority of the trades generated at the signal threshold of 50 or more continued on to achieve their 100% target. The neural signal strength continued to increase also in their research.

Monica Adya and Fred Collopy [11] found that nineteen studies (86%) produced the results favorable to forecasting and prediction through neural network (NN). In those papers NN outperformed alternative approaches.

Mr. S. P. Deshpande, H.V.P.Mandal Amravati and Dr. V. M. Thakare [12] have discussed about the applications of data mining in various field relevant to decision making such as medical science for diagnosis, sports world for player selection, game strategy etc., prediction of personal bankruptcy, for improving product demand forecasting, assorted optimization, product recommendation, assortment comparison across retailers and manufacturers, forecasting production schedules for the manufacturing plants, determining market potential in critical go/no decisions on continuing work, or making financial projections for stock holders and investors.

Xianjun Ni [13] has explained the various neural network methods in data mining. Author has discussed about data mining process based on neural network (NN). He has focused on applying various data mining types based on neural network (NN). Finally, he has explained about key techniques and approaches of implementation regarding data mining based on NN.

Kunwar Singh Vaisla and Dr. Ashutosh Kumar Bhatt [14] proved that neural network (NN) outperform statistical technique in forecasting stock market prices. They have showed it through a method to forecast the daily stock price using neural network and then the result of the neural network forecast is compared with

the Statistical forecasting result. They have proved that neural network, when trained with sufficient data, proper inputs and with proper architecture, can predict the stock market prices very well. On the other hand, statistical technique though well built but their forecasting ability is reduced as the series become complex. Therefore, NN can be used as a better alternative technique for forecasting the daily stock market prices.

Dase R.K. and Pawar D.D. [15] tried to sum up the application of Artificial Neural Network for predicting stock market. As per the authors view, in the World Predicting stock market index is a difficult task, but artificial neural network is having ability to predict stock index. They have also included that predicting stock index with traditional time series analysis has proven to be difficult and Artificial Neural network may be suitable for the task. A neural network (NN) has the ability to extract useful information from large set of data. They have presented a review of literature about application of artificial neural network (ANN) for stock market predictions and from this literature they have found that ANN is very useful for predicting world stock markets.

Muhammad A. Razi and Kuriakose Athappilly [16] have proved that NNs and CART models produce better prediction accuracy than non-linear regression model. They have also claimed that it is obvious from the study results that NNs and CART models provide better prediction compared to regression models when the predictor variables are binary or categorical and the dependent variable continuous. However they have also added that neither NNs nor CART model showed clear advantage of one over the other.

Chi-Jie Lu [17] has proved from the experimental results that the integrated independent component analysis (ICA)-based de-noising scheme with neural network proposed for stock price prediction model outperforms the integrated wavelet de-noising technique with BPN model, the BPN model with non-filtered forecasting variables, and a random walk model. According to the experiments, the author has concluded that the proposed method can effectively detect and remove the noise from stock prices/indices and improve the forecasting performance of BPN.

Pratyoosh Rai and Kajal Rai [18] have found from the comparison that problem of stock index prediction is one of the most popular targets for various prediction methods in the area of finance and economics. In Their article the researchers have described the comparison of different neural network types for stock prediction. The prediction was carried out by modular neural network, ARIMA-based neural network, Genetic algorithm, Amnestic neural network, Multi-Branch neural network etc. The authors have also performed comparative analysis of all these types of neural network (NN).

Debashish Das and Mohammad Shorif Uddin [7] In this paper, The researchers reviewing the possibility of applying two well-known techniques neural network and data mining in stock market prediction. As neural network is able to extract useful information from a huge data set and data mining is also able to predict future trends and behaviours. Therefore, a combination of both these techniques could make the prediction much reliable.

Conclusion

Based on methodological review, non-statistical techniques include measures that are related with the data mining and various neural network techniques such as genetic algorithm, multi-branch neural network may prove to be good performing methods for correct predictions if the model can be trained with sufficient data, proper inputs and architecture.

On the basis of the current survey analysis, we have found that NN method with expert system is the best performing method. Hence, the stock prediction system using a hybridization of knowledge-based data mining and neural network techniques would be a promising solution. Sometimes seen that association rule not applied but it can improve the stock market prediction broadly so that we will use NN method with association rule mining and pattern recognition technique to improve the prediction rate that will be useful for the share holders to analyse market to get a big profit. In future, it can be further investigated with real time stock data from any market.

Notes and References

[1]Govt. Engg. College, Banda
E-mail: nirmal7766@gmail.com

[1] Wang, Y.F., (2003) "Mining stock price using fuzzy rough set system", Expert Systems with Applications, 24, pp. 13-23.

[2] Wu, M.C., Lin, S.Y., and Lin, C.H., (2006) "An effective application of decision tree to   stock trading", Expert Systems with Applications, 31, pp. 270-274.

[3] Al-Debie, M., Walker, M. (1999). "Fundamental information analysis: An extension and UK evidence", Journal of Accounting Research, 31(3), pp. 261–280.

[4] Lev, B., Thiagarajan, R. (1993). "Fundamental information analysis", Journal of Accounting Research, 31(2), 190–215.

[5] Tsang, P.M., Kwok, P., Choy, S.O., Kwan, R., Ng, S.C., Mak, J., Tsang, J., Koong, K., and Wong, T.L. (2007) "Design and implementation of NN5 for Hong Kong stock price forecasting", Engineering Applications of Artificial Intelligence, 20, pp. 453-461.

[6] Ritchie, J.C., (1996) Fundamental Analysis: a Backto-the-Basics Investment Guide to Selecting Quality Stocks. Irwin Professional Publishing.

[7] Debashish Das and Mohammad Shorif Uddin (2013) "data mining and neural network techniques in stock market prediction" *International Journal of Artificial Intelligence & Applications* (IJAIA), Vol.4, No.1.

[9] Marijana Zekic. MS. "Neural Network Applications in Stock Market Predictions-A Methodology Analysis".

[10] Bruce J. Vanstone, Gavin Finnie & Clarence Tan.(2004), "Applying Fundamental Analysis and Neural Networks in the Australian Stockmarket", Bond University ePublications@bond,.

[11] Monica Adya & Fred Collopy.(1998), "How Effective are Neural Networks at Forecasting and Prediction? A Review and Evaluation", *Journal of Forecasting, J. Forecast*, 17, 481 – 495.

[12] S. P. Deshpande & V. M. Thakare.(2010), "data mining system and applications: a review", *International Journal of Distributed and Parallel systems* (IJDPS), Vol.1, No.1.

[13] Xianjun Ni.(2008), "Research of Data Mining Based on Neural Networks", World Academy of Science, Engineering and Technology, 39.

International Journal of Artificial Intelligence & Applications (IJAIA), Vol.4, No.1, January 2013

[14] Kunwar Singh Vaisla & Dr. Ashutosh Kumar Bhatt.(2010), "An Analysis of the Performance of Artificial Neural Network Technique for Stock Market Forecasting", (IJCSE*) International Journal on Computer Science and Engineering*, Vol. 02, No. 06, pp. 2104-2109.

[15] Dase R.K. & Pawar D.D.(2010), "Application of Artificial Neural Network for stock market predictions: A review of literature*",

International Journal of Machine Intelligence, ISSN: 0975–2927, Volume 2, Issue 2, pp. 14-17.

[16]   Muhammad A. Razi & Kuriakose Athappilly.(2005), "A comparative predictive analysis of neural networks(NNs), nonlinear regression and classification and regression tree (CART) models", Expert Systems with Applications, 29, pp. 65–74.

[17] Chi-Jie Lu, "Integrating independent component analysis-based denoising scheme with neural network for stock price prediction," Expert Systems with Applications 37(10), pp. 7056-7064.

[18] Pratyoosh Rai & Kajal Rai.(2011), "Comparison of Stock Prediction Using Different Neural Network Types", *International Journal of Advanced Engineering & Application.*