

Stock Market Prediction Approach: An Analysis

Nikita Agrawal
Software Engineer
HSBC Technology India
Pune, India

Abstract:- The data mining is the approach which can extract useful information from the large amount of data. The prediction analysis is the approach of data mining which can predict future possibilities based on current information. The prediction analysis approach has various techniques like pre-processing, feature extraction and classification. This review paper is based on the stock market prediction. In this paper, various stock market prediction techniques are reviewed which are based on classification. The stock market prediction techniques are implemented in python

Keywords:- Stock market, machine learning, feature extraction

INTRODUCTION

Data mining can be described as “making better use of data”. Every human being is increasingly faced with unmanageable amounts of data; hence, data mining or knowledge discovery apparently affects all of us. It is therefore recognized as one of the key research areas. Ideally, we would like to develop techniques for “making better use of any kind of data for any purpose”. However, we argue that this goal is too demanding yet [1]. Over the last three decades, increasingly large amounts of historical data have been stored electronically and this volume is expected to continue to grow considerably in the future. Yet despite this wealth of data, many fund managers have been unable to fully capitalize on their value. Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area [2]. 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. Public companies raise funds by issuing shares to public and institutions. By purchasing shares, an investor acquires partial ownership of company. A stock market is a place where public listed company stocks are traded. Stocks are exchanged among buyer and seller which generate transaction data. Prices are changing as per demand and supply of stocks [3]. All trading data is captured by stock exchange where stock companies are listed. Stock trading data is nonlinear, fluctuating hence highly time variant. A lot of information is hiding in this huge data captured by stock exchange is difficult and time consuming for human being to extract without powerful tools. Thus predicting future price of stock is highly

challenging. Stock prices are changing due to demand and supply of stocks in stock market. Based on past history and current information, investors are buying or selling stocks of listed companies. Investors, stock analyst and stock brokers are predicting demand and supply of stocks after studying fundamentals and technical information of stocks [4]. In order to study and analyze stocks data, investor need to go through vast amount of data to discover hidden patterns which is very cumbersome and tiring task. Usually stock analysts analyze stocks based on three vital aspects - external environmental factors, fundamental analysis and technical analysis. There are several theories applied for stock market prediction among which some are explained below:

i. Efficient Market Hypothesis (EMH): Efficient Market Hypothesis was an idea developed in the 1965. EMH states that the price of a security will reflect the whole market information. As soon as there is any information indicating that a security is underpriced and therefore offers a profit opportunity, investors will immediately buy it and its price will rise up to a fair value. There were three forms of financial market efficiency introduced: weak, semi-strong and strong [5]. In weak-form efficiency, future prices cannot be predicted by analyzing prices from the past; hence the technical analysis was unreliable. In semi-strong-form efficiency, the price will rapidly adjust to publicly available new information; neither fundamental analysis nor technical analysis techniques are able to reliably produce excess returns.

ii. Behavioral finance: Behavioral finance explains why and how markets might be inefficient. It is to study the influence of psychological, social, cognitive, and emotional factors on the security and the markets. The historical patterns and the current market circumstance can psychologically affect the investor, and their respond will lead to a predictable market trend [6].

Some of the important methodologies of stock market prediction include the following techniques:

a. Time Series Forecasting (TSF): Time series forecasting is a process of analyzing time series data and predicting the future outcome. It uses statistical techniques to model and explain time-dependent series of data points. TSF is a widely used technique in financial forecasting.

b. Back Propagation Neural Network (BPNN): Neural network (NN) is a computational model that is capable of estimation (TSF) and pattern recognition (Classification). It is robust with respect to noisy and erroneous data, and is able to learn and adapt to the environment. Neural network is applicable to the problem that algorithm is indefinable or exhaustive search is infeasible. NN is in consideration as

long as the optimal solution is the first priori, not the interpretable result. In this work, the neural network is multi-layer perceptron (MLP) and uses a back propagation (BP) algorithm in its learning.

Classification consists of predicting a certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute [7]. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. Next the algorithm is given a data set not seen before, called prediction set, which contains the same set of attributes, except for the prediction attribute – not yet known. The algorithm analyses the input and produces a prediction. The prediction accuracy defines how “good” the algorithm is.

i. Decision tree (D-Tree): Decision tree is a classification method which yields output as a flowchart-like tree structure. The result from D-Tree is highly interpretable, but the outcome must be represented in categorical data [8]. Mostly, DTree algorithm called “J48” is applied to classify future stock market direction.

ii. Naïve bayes: Naïve Bayes is a simple probabilistic classifier based on Bayes’ theorem, with a naive assumption of independence between every pair of features.

iii. k-Nearest neighbors (KNN):KNN is a non-parametric lazy learning algorithm that predicts class of the object based on the k closest training examples in the feature space. An object is classified by a majority vote of its neighbors; the object will be assigned to the class most common amongst its k nearest neighbors (k = 15 in this study).

LITERATURE REVIEW

Ze Zhang, et.al (2017) proposed the utilization of Elman network to predict the opening price of stock market. Typically, Elman neural network is a local recurrent neural network, having one context layer that memorizes the past states, which is quite fit for resolving time series issues. Considering that Elman network is limited, this paper adopts self-adapting variant PSO algorithm to optimize the weights and thresholds of network [9]. Finally, this paper verifies that model by some stock prices, and compares with BP network and Elman network, so as to draw the result that shows the precision and stability of this predication model both are superior to the traditional neural network.

Nonita Sharma, et.al (2017) emphasized on the prediction of future stock market index values based on historical data. The experimental evaluation is based on historical data of 10 years of two indices, namely, CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex from Indian stock markets [10]. The prediction performance of the proposed model is compared with that of well-known Support Vector Regression. Technical indicators are selected as inputs to each of the prediction models. The closing value of the stock price is the predicted variable. Results show that the proposed scheme outperforms Support Vector Regression and can be applied successfully for building predictive models for stock prices prediction.

Yaojun Wang, et.al (2016) used the social media mining technology to quantitative evaluation market segment, and

in combination with other factors to predict the stock price trend in short term. In this paper, starting from the efficient market hypothesis the authors fetch the stock comments information from social media and then preprocessing the data to emotion vectors [11]. In addition, through testing our social media mining algorithm we discovered that the SVM model contains segment index has higher prediction accuracy than SVM model not combined segment index. Experiment results show that by using social media mining combined with other information, the stock prices prediction model can forecast more accurate.

Mustain Billah, et.al (2016) proposed an improved Levenberg Marquardt (LM) training algorithm of artificial neural network [12]. Improved Levenberg Marquardt algorithm of neural network can predict the possible day-end closing stock price with less memory and time needed, provided previous historical stock market data of Dhaka Stock Exchange such as opening price, highest price, lowest price, total share traded. This improved LM training algorithm proves neural network to be better computing tool for predicting closing stock price in Bangladesh Stock Exchange perspective. Moreover, it can be used for any prediction purposes. In future, we will use this concept in predicting network traffic prediction purpose.

R. Yamini Nivetha, et.al (2017) analyzed various prediction algorithms to build a prediction model. The prediction model will be based on monthly prediction and daily prediction to forecast the next day market price [13]. This model estimates the open value of the next day in the market. A comparative study of these three algorithms which are Multiple Linear Regression, Support Vector Machine and Artificial Neural Network are done. The stock price is predicted by sentiment analysis with the best forecasting algorithm. The result exhibits that the deep learning algorithm performs better than the MLR and SVM. In deep learning algorithm the hidden layer neuron learns in every prediction. Hence the output layer neuron produces the best outcome. Artificial Neural Network is the best predicting algorithm.

Harun Ercan, (2017) presented that in Baltic countries; artificial neural networks are not commonly used in predicting financial failures. This study aims using artificial neural networks to predict OMX Baltic Benchmark GI (OMXBBGI) market index value [14]. The research performed on the Baltic stock market uses index value and EUR/USD exchange rate to predict the index. The training set is determined to include about 70% of the data set, and 15% will be used for testing, and the rest of 15% will be used for validation purposes. MATLAB ANN tool is used to complete the tests. The results showed that Baltic Market values are successfully predicted by NARX methodology with the variables: Exchange rate and previous days’ index values. The uniqueness of the paper especially comes from variables that have never been tried to forecast Baltic Stock Market values by ANN.

Muhammad Waqar, et.al (2017) investigated the problem of high dimensionality of stock exchange to predict the market trends by applying the principal component analysis (PCA) with linear regression [15]. Experiments are carried out on a high dimensional spectral of 3 stock exchanges such

as: New York Stock Exchange, London Stock Exchange and Karachi Stock Exchange. The accuracy of linear regression classification model is compared before and after applying PCA. The experiments show that PCA can improve the performance of machine learning in general if and only if

relative correlation among input features is investigated and careful selection is done while choosing principal components. Root mean square error (RMSE) is used as an evaluation metric to evaluate the classification model.

TABLE 1: TABLE OF COMPARISON

Authors Names	Year	Description	Outcomes
Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu	2017	Considering that Elman network is limited, this paper adopts self-adapting variant PSO algorithm to optimize the weights and thresholds of network.	This paper verifies that model by some stock prices, and compares with BP network and Elman network, so as to draw the result that shows the precision and stability of this predication model both are superior to the traditional neural network.
Nonita Sharma, Akanksha Juneja	2017	The prediction performance of the proposed model is compared with that of well-known Support Vector Regression.	Results show that the proposed scheme outperforms Support Vector Regression and can be applied successfully for building predictive models for stock prices prediction.
Yaojun Wang	2016	In this paper, starting from the efficient market hypothesis the authors fetch the stock comments information from social media and then preprocessing the data to emotion vectors.	Experiment results show that by using social media mining combined with other information, the stock prices prediction model can forecast more accurate.
Mustain Billah, Sajjad Waheed, Abu Hanifa	2016	An improved Levenberg Marquardt (LM) training algorithm of artificial neural network was proposed.	This improved LM training algorithm proves neural network to be better computing tool for predicting closing stock price in Bangladesh Stock Exchange perspective.
R. Yamini Nivetha, Dr. C. Dhaya	2017	A comparative study of these three algorithms which are Multiple Linear Regression, Support Vector Machine and Artificial Neural Network are done.	Hence the output layer neuron produces the best outcome. Artificial Neural Network is the best predicting algorithm.
Harun Ercan	2017	This study aims using artificial neural networks to predict OMX Baltic Benchmark GI (OMXBBGI) market index value.	The uniqueness of the paper especially comes from variables that have never been tried to forecast Baltic Stock Market values by ANN.
Muhammad Waqar, Hassan Dawood, Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar, Ping Guo	2017	The problem of high dimensionality of stock exchange was investigated to predict the market trends by applying the principal component analysis (PCA) with linear regression.	Root mean square error (RMSE) is used as an evaluation metric to evaluate the classification model.

CONCLUSION

In this work, it is concluded that stock market prediction is the major issue of the prediction analysis due to high complexity of the dataset. The stock market prediction approach has various steps like feature extraction and classification. The stock market prediction techniques which are designed so far are based on the classification method. In future hybrid classification method will be designed which give high accuracy for the stock market prediction

REFERENCES

- [1] S. Qamar, & S. H. Adil, "Comparative analysis of data mining techniques for financial data using parallel processing," In Proceedings of the 7th International Conference on Frontiers of Information Technology, 2009.
- [2] S. H. Adil, & S. Qamar. "Implementation of association rule mining using CUDA," International Conference on Emerging Technologies (ICET), 2009.
- [3] S. S. A. Ali, M. Moinuddin, K. Raza, & S. H. Adil, "An adaptive learning rate for RBFNN using time-domain feedback analysis," The Scientific World Journal, 2014.
- [4] S. Asur, & B. A. Huberman, "Predicting the future with social media," International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010.
- [5] C. D. R. Aurangzeb, "Factors Affecting Performance of Stock Market: Evidence from South Asian Countries," International journal of academic research in business and social sciences, vol. 2, no. 9, 2012.
- [6] S. Shen, H. Jiang, & T. Zhang, "Stock market forecasting using machine learning algorithms, 2012.
- [7] I. Ali, K. U. Rehman, A. K. Yilmaz, M. A. Khan, & H. Afzal, "Causal relationship between macro-economic indicators and stock exchange prices in Pakistan," African Journal of Business Management, vol. 4. No. 3, pp. 312, 2010.
- [8] J. Bollen, H. Mao, & X. Zeng, "Twitter mood predicts the stock market. Journal of Computational Science," vol. 2, no. 1, pp. 1-8, 2011.
- [9] Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu, "Short-term Prediction for Opening Price of Stock Market Based on Self-adapting Variant PSO-Elman Neural Network", 2017, IEEE
- [10] Nonita Sharma, Akanksha Juneja, "Combining of Random Forest Estimates using LSboost for Stock Market Index Prediction", 2017 2nd International Conference for Convergence in Technology (I2CT)
- [11] Yaojun Wang, Yaoqing Wang, "Using Social Media Mining Technology to Assist in Price Prediction of Stock Market", 2016 IEEE International Conference on Big Data Analysis (ICBDA)
- [12] Mustain Billah, Sajjad Waheed, Abu Hanifa, "Stock Market Prediction Using an Improved Training Algorithm of Neural Network", 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) 8-10 December 2016
- [13] R. Yamini Nivetha, Dr. C. Dhaya, "Developing a Prediction Model for Stock Analysis", 2017 International Conference on Technical Advancements in Computers and Communications
- [14] Harun Ercan, "Baltic Stock Market Prediction by Using NARX", 2017, 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)
- [15] Muhammad Waqar, Hassan Dawood, Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar, Ping Guo, "Prediction of Stock Market by Principal Component Analysis", 2017 13th International Conference on Computational Intelligence and Security