

Survey on Stock Market Prediction and Performance Analysis

Mrityunjay Sharma

Abstract— Predicting stock market accurately has always intrigued the market analysts. During the past few decades various machine learning techniques have been applied to study the highly stochastic nature of stock market by capturing and using repetitive patterns. Different companies use different analysis tools for forecasting and the main aim is the accuracy with which they predict which set of stocks would yield the maximum amount of profit. This paper gives a brief introduction to various techniques used for prediction so that it is easy for buyer/seller to decide.

Keywords— Neural network, Regression, Time series model.

I. INTRODUCTION

Time series involves studying the observations chronologically at regular intervals. Global crashes are headed by local crashes and do not happen all of a sudden. The increase in the use of temporal data has initiated various development and research efforts in the area of data mining. Time series is an important class of temporal data observations and can be easily obtained from scientific and financial applications (e.g. weekly sales totals, daily climate prediction, prices of mutual funds and stocks). Time series is an important high-dimensional data type. Unlike static data, the time series of a feature comprise values changed with time. Time series data are of interest because of its pervasiveness in various areas ranging from science, engineering, meteorology, business, stock market, economic, health care, to government. Time series is predicts future observations based on the some historic observations noted [1].

The main focus of machine learning is on prediction, based on previously known properties that are acquired and learnt from training data. It makes an effort to completely remove the need for human instinctive knowledge in the field of data analysis. It can be used in recognizing a handwritten or printed character, predicting climate and a vast variety of applications. Various methods are being used for this purpose such as regression, support vector machine, genetic algorithms, clustering etc. An attempt is also being made to

Manuscript received Jan, 2014.

Mrityunjay Sharma, Department of Computer Science, Jaypee University of Information Technology, Waknaghat, Solan, India, Mobile No:- +91-9418371963

combine all these approaches and increase the predictive power of the model proposed by also reducing the memory requirement for the samples by keeping only relevant observations.

One of the vital source for companies to increase money is stock market. By distributing shares of ownership of the company in a public market raises supplementary financial capital for expansion and permits businesses to be traded publicly by selling shares of ownership of the company in a public market. Stock market can effect the social mood to a great extend. The value of the stock and other assets is considered as an important parameter how we can judge the dynamics in the field of economic movement. An economy where the stock market is on the hike is considered to be an up coming economy. Rising prices of shares is generally linked with increased business investment and vice versa. It is very important to predict the stock as it may lead to high profit or huge loses that the buyer/seller has to bear. The prediction that is done helps in decision making process whether to buy or sell a share. Some techniques such as the Kappa measure also allow making decisions based on considering the various techniques and the results obtained from them [2].

The remainder of the paper is organized as follows, In section II, we discuss about the Artificial neural network. In section III and IV, we discuss about the Regression and Support Vector Machine. In section V, we briefly review the related work. In section VI, we describe the comparison of the techniques. Finally in section VII, we summarize our conclusion.

II. ARTIFICIAL NEURAL NETWORK

ANN is a mathematical model that has been inspired by the animal nervous system consisting of neurons and the way information is traveled from every part of human body to the brain. Information in the form of some data values is fed into the network (interconnection between neurons). Based on specific function used at each layer and the input value the output can be evaluated. It has three types of layers:-

- (a) The input layer receives the values on which the computation has to be done. These are the different values of the tuples in the dataset.
- (b) In the hidden layer the computation are done as the values are passed through each level. The number of hidden layers may vary in different architectures and applications.
- (c) In the output layer we receive the value of the parameter after have been processed and computed by specific

activation function according to the application for which it is designed. It may be a numeric, binary or a categorical value.

ANN has found its application in classification, robotics, regression, time series evaluation and many more. Learning is done by calculating the mean square error for each subsequent observation and a model is chosen that has the least error and high predictive power.

III. REGRESSION

Regression is a statistical tool which is used in order to study connection or the relationship between an independent variable or regressor variable (such as GDP, population, education) and the dependent variable or the response variable (such as the employment). This means that the employment rate can be calculated as a function of GDP, population and educational status. It can be linear or non

linear. Recent researches have seen the combination of other machine learning techniques such as nearest neighbor, genetic algorithm and support vector with regression. [3]–[5]. Various methods are combined in order to increase predictive power of the model, reduce the number of variables involved. Value of variables may sometimes come from intensive experiments and data collection so reducing variables can help reduction in this effort also. In order to estimate the value of the variable accurately the variables are sometimes transformed. Stepwise selection, Backward elimination and forward selection are the basic methods used to build a best regressor model i.e. building a model that has only the explanatory variables that have high significance in the model and less multicollinearity (linear relationship between variables). This leads to reduction in the number of independent variables that are to be considered. General equation for multiple linear regression having n vector of regressor is

$$Y = X\beta + \epsilon$$

where Y is matrix of dependent variable, X is matrix of independent variables, β is the regression coefficients and ϵ is the random error that might be introduced in the model.

IV. SUPPORT VECTOR MACHINE

Vladimir N. Vapnik invented the original support vector algorithm. This concept has been used for studying the linear and non-linear classification by analyzing the data and observing the patterns that are observed. It falls under the category of supervised machine learning. Different hyperplanes may be constructed in order to classify data. Our main aim is to select the maximum marginal hyperplane which is the plane that has the maximum distance between each other hence making a robust model. It is not desirable to have data points on the plane (support vectors) that classify the data. A data point is represented as a vector (a list of p numbers) and having p dimensions and a decision has to be made whether we can separate all such points with a hyperplane having $(p-1)$ -dimension. This is called a linear

classifier. Support vector technique is used in a variety of applications ranging from the classification of proteins to recognizing hand written characters.

V. METAHEURISTIC ALGORITHMS

These algorithms are not problem specific and are non deterministic in nature. Recently there has been a tremendous use of genetic algorithms, evolutionary computing and swarm intelligence in the field of optimization and have been applied to various fields ranging from resource sharing in cloud computing to mathematical function optimization. Global optimum solution is not guaranteed for certain class of problems. It has been observed that it can search large search space with less computational efforts.

VI. RELATED WORK

A. Neural Network Time Series Model

Artificial neural networks are generally considered to have high predictive power but the main problem associated with it is that some very small changes in the training data that is used for applying the machine learning algorithms can lead to large changes in the prediction. Thus they have many disadvantages when applied to time series model. Stock exchange which is a highly stochastic and a dynamic activity is highly affected by the macro and micro economical activities that occur. Natural measure of performance that is used is the prediction error. As consideration of one error is not desirable to predict a system so Mean absolute error, mean squared error, mean absolute percentage error, root mean squared error and percentage mean absolute deviation is used [6].

Time series model using neural network uses the back propagation feed forward network with one input, output and hidden layer each. Predicting the closing price one day in advance is done by it. It uses the Gradient descent algorithm as the learning function and the sigmoid function as the activation function. It has been observed while using this approach for predicting NIFTY and MIDCAP50 that it performs well when the dataset used has data of 1 to 5 years. This model works well when the noise is less and prediction accuracy is reduced considerably when the noise is increased.

B. Fundamental Analysis Based Method

Fundamental Analysis based approach is developed and designed by studying a variety of macroeconomics, industrial environment, financial condition and financial news [7]. This approach improves the decision making capability of the investor and helps him to invest wisely, earning maximum profit. This approach is performed in a series of steps.

a) Weight Calculation of Financial Indicator - From the Taiwan Economic Journal (TEJ) the financial indicators for each individual stock firm is captured after which they are normalized. Weights of the normalized financial indicator is calculated by analyzing the correlation between the stock and financial indicator by employing the GRA (Grey relational analysis) [8].

b) Individual Stock Evaluation and Selection - CCR model [9] is used to calculate the technical efficiency of the technical indicator after which they are arranged in the descending order and the investor can select the one that has the maximum value of stock and predict its future price.

c) Feature Selection of Financial News - News can also have a significant effect on the prices of stock. By considering the historic news, analysis and capturing the words related to stock a database of features relating to financial news is constructed. Naive Bayesian classifier is used with financial news as the input and the effect of these news on the stock trading as the output.

d) Stock Price Trend Forecasting - This process involves the retrieval of indicators of stock by financial news and the TEJ and Taiwan Stock Exchange Corporation which gives indicators relating to macroeconomics and industry. Strong classifier by using the queen genetic algorithm and this is used to trading signal of stock as "Down", "Flat" or "Up".

C. Stock Trend Prediction by Using Regression Analysis

Linear regression in conjunction with the least square method for determination of parameters and moving average method is used to predict the stock market time series data [10]. The moving average method is performed by taking the average of value of observation of a particular time span and placing the value in between the span. The reason for selecting this method is that it tends to reduce the fluctuation in market and obtain the value of the trend of time series with high accuracy. The percentage earning which is the ratio of current market price and the earning per share is used to make the regression equation. The system is able to predict the monthly and weekly dynamic movement of stock prices. The main benefit of this approach is that it yields a numeric forecasting method which has advantage and is easy to understand in comparison to some words such as high, low, medium, flat, which may be ambiguous if their range is not mentioned.

D. Support Vector Machine for Stock Prediction

Support vector can be applied to forecast the stock exchange by using some company specific parameters such as price per earnings ratio of stock, net income, net revenue, diluted earnings per share. Though it is difficult and complex to define whether a market is good or not this method defines good when the share of a particular company rose over a year else termed bad. It is also difficult to determine the parameters that should be considered for checking whether value of share rose. SVM is used for

classification and can be used for it. SVM is considered good for classifying non linear patterns by developing a concrete relationship between the class labels and input variables. It has been observed that the model proposed can have higher accuracy by adding additional variables. Out of the 10 samples that were to be predicted it yielded the correct result for [7].

E. Optimal Neural Network Architecture

By focusing only on the market close value and the open value of a share this architecture advises the user to either withhold, buy or sell a share resulting in good decision capability [12]. Data is obtained from the yahoo Finance and it is represented as a time series model. The ANN that is used consists of 1 input, 2 hidden layer and 1 output layer and considers maximum epochs to be 5000. By empirical testing it is found that the optimum number of neuron in the first hidden layer should be half of that in the input layer with the input layer consisting of 44 neurons. The mean of both the actual and predicted data is calculated and there is a situation where the mean of data predicted is more than the actual data obtained from the training set then the system suggests the investor to buy a particular stock or else he is suggested not to buy the stock and wait for appropriate time to buy the same.

F. Adaptive Linear Combiner (ALC) and Artificial Bee Colony Algorithm (ABC)

Artificial Bee Colony is a meta heuristic algorithm inspired from the behaviour of the bee. It consists of various phases and at each phase we select a better solution or in other term a better value of fitness function. This can be understood as bees visiting sources that have greater nectar than the other patches of flowers in the neighborhood. With a major effort to make adaptive and decentralized algorithms swarm intelligence is being evolved as a major area of research at present. Studying the behaviour of various organisms like the honey bee, ant colony and E. coli bacteria helps in the global optimization of a certain application specific function. This model can predict short and long term stock prices [13]. The adaptive linear combiner (ALC) is used in the model and the weights are then trained thereafter by applying the ABC algorithm. Mean absolute percentage error is used to find an optimum solution. When compared with genetic algorithm and particle swarm optimization this model performs extremely well giving high accuracy as compared to them.

S no.	Technique	Advantages	Disadvantage	Parameter used
A	Neural Network Time Series Model	1)Performance better than regression. 2)lower prediction error	Prediction gets worse when noise variation is increased	Closing price of stock
B	Fundamental Analysis Based Method	1)Forecast to select optimum stock and predict trend. 2)Output Up, Flat or Down and not numeric.	Feature selection from financial news complex and typical	Factors of macroeconomics, financial news, industrial environment and firm financial condition.
C	Stock Trend Prediction by Using Regression Analysis	1)Numeric forecasting method. 2)Predict six month future value 3)Moving average reduces the fluctuations.		Percentage earning ratio calculated from daily and weekly activities whose data extracted from stock exchange
D	Support Vector Machine for Stock Prediction	Does not lose much accuracy when applied to a sample from outside the training sample.	Can exaggerate minor fluctuations in the training data, thus resulting in decrease in subsequent predictive ability.	net revenue, net income, price per earnings ratio of stock, consumer spending, diluted earnings per share, unemployment rate, consumer investment, federal funds rate, dow jones industrial average and inflation rate
E	Optimal Neural Network Architecture	Select the optimized parameters in terms of epochs and neurons to provide accurate forecasting.	Only binary classification for withholding or buying the stock in the market.	Earning per share, price dividend ratio and price earning ratio.
F	Adaptive Linear Combiner (ALC) and Artificial Bee Colony Algorithm(ABC)	Better than other two adaptive parameter learning algorithms particle swarm optimization and genetic algorithm with the adaptive linear combiner.		There are 10 input to model. Performance evaluated by mean absolute percentage error(MAPE).

Table I- COMPARISON OF ALL THE TECHNIQUES

ABOUT THE AUTHOR

CONCLUSION

Stock exchange prediction help the organization and also the stake holder to keep track of the trend of the market. It also helps to decide whether to sell, buy or withheld the stock so as to maximize the profit. In this paper, I have made a comparative study of various techniques that are used to predict the stock market giving a brief description of each. Discussions of those techniques are reviewed and the benefits and drawbacks of feature various techniques explained are summarized.

REFERENCES

- [1] S. Makridakis, "Forecasting: its role and value for planning and strategy," *International Journal of Forecasting*, vol. 12, no. 4, pp. 513–537, 1996.
- [2] R. Gupta, N. Garg, and S. Singh, "Stock market prediction accuracy analysis using kappa measure," in *Communication Systems and Network Technologies (CSNT), 2013 International Conference on*, 2013, pp. 635–639.
- [3] A. Widodo, M. Fanani, and I. Budi, "Enriching time series datasets using nonparametric kernel regression to improve forecasting accuracy," in *Advanced Computer Science and Information System (ICACSIS), 2011 International Conference on*, 2011, pp. 227–232.
- [4] L. You and Y. Lu, "A genetic algorithm for the time-aware regression testing reduction problem," in *Natural Computation (ICNC), 2012 Eighth International Conference on*, 2012, pp. 596–599.
- [5] L. Lorenzi, G. Mercier, and F. Melgani, "Support vector regression with kernel combination for missing data reconstruction," *Geoscience and Remote Sensing Letters, IEEE*, vol. 10, no. 2, pp. 367–371, 2013.
- [6] D. A. Kumar and S. Murugan, "Performance analysis of indian stock market index using neural network time series model," in *Pattern Recognition, Informatics and Medical Engineering (PRIME), 2013 International Conference on. IEEE, 2013*, pp. 72–78.
- [7] Y.-J. Chen and Y.-M. Chen, "A fundamental analysis-based method for stock market forecasting," in *Intelligent Control and Information Processing (ICICIP), 2013 Fourth International Conference on*, 2013, pp. 354–359.
- [8] X. Zhou, W. Jiang, Y. Tian, and Y. Shi, "Kernel subclass convex hull sample selection method for svm on face recognition," *Neurocomputing*, vol. 73, no. 10, pp. 2234–2246, 2010.
- [9] A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European journal of operational research*, vol. 2, no. 6, pp. 429–444, 1978.
- [10] S. A. S. Olaniyi, K. S. Adewole, and R. Jimoh, "Stock trend prediction using regression analysis—a data mining approach," vol. 1, pp. 154–157, 2010.
- [11] Z. Hu, J. Zhu, and K. Tse, "Stocks market prediction using support vector machine," in *Information Management, Innovation Management and Industrial Engineering (ICIII), 2013 6th International Conference on*, vol. 2, 2013, pp. 115–118.
- [12] G. Khirbat, R. Gupta, and S. Singh, "Optimal neural network architecture for stock market forecasting," in *Communication Systems and Network Technologies (CSNT), 2013 International Conference on*, 2013, pp. 557–561.
- [13] M. Rout, U. M. Mohapatra, B. Majhi, and R. Mahapatra, "An artificial bee colony algorithm based efficient prediction model for stock market indices," in *Information and Communication Technologies (WICT), 2012 World Congress on. IEEE, 2012*, pp. 750–754.



Mrityunjay Sharma received B.Tech from Himachal Pradesh University, Shimla (H.P) India in Information Technology in the year 2011. He is currently working as Teaching assistant and is pursuing M.Tech. Degree in Computer Science and Engineering from Department of Computer Science, Jaypee University of Information Technology, Waknaghat, Solan, HP (India). His research area is Machine Learning and its applications.