Stock Market Index Forecasting of Nifty 50 Using Machine Learning Techniques with ANN Approach

Gourav Kumar
Department of computer Science and IT
University of Jammu
Jammu, India
E-mail: gourav4706@gmail.com

Vinod Sharma
Department of computer Science and IT
University of Jammu
Jammu, India
E-mail: vnodsharma@jammuuniversity.in

Abstract: Stock market index forecasting is one of the most challenging tasks for one who wants to invest in stock market. This challenge is because of the uncertainty and volatility of the stock prices in the market. Due to advancement in technology and globalization of business and share markets it is important to forecast the stock prices more quickly and accurately. Machine Learning techniques are most accepted due to the capability of identifying stock trend from massive amounts of data that capture the underlying stock price movement. Artificial Neural Network (ANN) algorithms are mostly implemented and play an important role in decision making for stock market index predictions. Multi Layer Perceptron (MLP) architecture with back propagation algorithm has the capability to predict with greater accuracy. This paper presents an ANN based approach to forecast Nifty 50 Index. A feed-forward neural network using multiple back propagation algorithm has been used to forecast next day’s OHLC data. This model has used the pre-processed dataset of Open price (O), High price (H), Low price (L), Close price (C), Volume Traded (V) and Turnover (T) for the period of 10 years from 03 April 2006 to 16 May 2016. The root mean square error (RMSE) is chosen as indicators of performance of the network. The performance of this model has been tested to an average accuracy of 99.2152 % with an RMSE error of 0.0079. In this research Multiple Back-Propagation (MBP version 2.2.4) software has been used to predict the future stock prices and their performance statistics have been evaluated. This would help the investor to take better business decisions such as buy or sell a stock.

Keywords: Machine Learning, Artificial Neural Network (ANN), Feed Forward Neural Network, Multi-Layer Perceptron (MLP), National Stock Exchange (NSE), Stock Market Prediction.

I. INTRODUCTION

A share market is a market in which shares of publicly held companies or people are traded to raise money. The fluctuation of prices of shares depends upon the demand and supplies of shares. Only the registered Companies are allowed to carry out trading. Stock market forecasting is the process of trying to predict the future stock value of a company. The successful prediction of a stock's prices would return significant profit. Forecasting stock market index is an important financial problem that is receiving increasing diligence. A number of neural network models have been proposed for attaining accurate prediction results. Economies of a country are strongly linked and heavily influenced by the performance of the stock market. In India stock prices are basically influenced by sentimental factors and event driven activities. The investors at a large are still at a nascent stage and lack required knowledge and experience before investing in stock market. Most of them just do it for the sake of trying their luck. A few who own good resources and are serious about it often tend to invest based on good fundamentals, technical indicators and news channel based advises. We often see a sudden spurt or spike in stock prices as a result of these. There are many techniques in the literature and applications to predict short term movements based on different stochastic models. Such approaches generally rely on treating individual stock data as a time series without analyzing correlations and patterns between different companies, mainly because of the limitations of processing very large data sets at very high-speeds. We have proposed here a model which will not require any previous experience of trading for investing into the stock market. The model suggests trend of stock index based on historical prices.

In general the stock market prediction is done using the following four methods:
A. Technical Analysis
B. Fundamental Analysis
C. Time Series Analysis
D. Machine Learning Methods

Technical Analysis is done by plotting technical data such as Open, High, Low and Close prices on charts and using them to detect trends. Trends are assumed to be based on supply and demand issues which are believed to be cyclic or in patterns. The problem here is that the extraction of trading rules from these charts is highly subjective in nature. Therefore, each analyst gives a different trading view by studying the same chart. A fundamental analyst on the other hand aim to compute the intrinsic value of an underlying stock, taking into account the variables such as the growth, dividend payout, the interest rates, investment, the sales level, the tax rates, P/E ratio, credit risk and so on. In traditional methods of stock prediction analysts generally try to learn the influences with regard to the past performance to predict the future with regression analysis. Regression models have been used to predict stock market time series. Stock prices are often more oriented to the future performance of a company and its sector. Also, the authors believe that the influence factors on performance do change a lot and gradually over time. With advent of latest technology and resources at every hand it is difficult to predict the market using age old traditional methods. The market is evolving new trends and paths every new day and to cater this fast pace of change latest techniques of machine learning and artificial neural network needs to be deployed for making a guess of market movements.[1]
II. RELATED WORK

In literature, different sets of input vectors are used for stock market prediction. Different input variables have been used to predict the same dataset of stock market. Some researchers used inputs from time series and some considered the heterogeneous market information and macroeconomic variables.

Lot of research work has been done and models based on a range of intelligent soft computing techniques are developed for stock market forecasting. This section describes briefly some of the work that has already been done in the field of stock market.

K.K Suresh kumar and Dr. N.M Elango [2] have used neural works predict tools to predict the future stock prices and evaluated their performance statistics. In this research, they studied and applied multilayer Perceptron model by using the NeuralWorks Predict tool. The results from the study shows that NeuralWorks Predict have a ability to predict the stock prices more accurately than the other existing tools and techniques.

Sukhdev Singh Chauhan, Vinod Sharma et al[3], have presented an ANN approach to forecast S&P Nifty index. In this research work feed-forward Neural Network (Multi-Layer Perceptron) with multiple back propagation algorithm have been used to predict the next day’s OHLC data. This model has used the pre-processed data of Open (O),High(H),Low(L),Close(C),Volume(V)and turnover(T) for the period of seven years from 02 Jan 2006 to 24 Oct 2013. The performance of this model have been tested to an average accuracy of 98.6187% with an RMSE of 0.01381.

Suganda Saha et al[4] considered feed forward neural network (FFNN) model for stock market prediction and compared its result with Radial basis function network (RBFN) model, fuzzy logic model and Elman network model. In this paper, Levenberg-Marquardt Back-propagation algorithms have been used to train the data for both FFNN and Elman network. For Fuzzy Logic, Sugeno type Fuzzy Inference System (FIS) is used to model the prediction process.

Prakash Ramani and Dr. P.D Murarka [5] have proposed a stock market prediction model using Multi-Layer feed-forward Neural Network. In this model they have used back-propagation algorithm. As Closing price of any stock covers all other attributes of the company. Hence they have used historical close prices for training the network. The testing was performed on ten different companies’ data and the results obtained were quite satisfactory.

Zabir Haider Khan et al[6] used Back-propagation algorithm and Multilayer Feed-forward network for predicting share market price. This model has used the pre-processed data of five parameters for training the network. These parameters are: General Index (GI), P/E ratio, Net Asset Value (NAV), Earning Per Share (EPS) and Volume. They used feed-forward neural network which had a input layer with five nodes, a hidden layer which has five nodes and a output layer with one node. They inferred that when they consider three inputs for prediction the sum squared error was not satisfactory. But when they consider four inputs the sum squared error was minimize.

Emad W. Saad et al[7] performed comparative study of stock trend prediction using Time Delay Neural Network(TDNN), Recurrent Neural Network(RNN) and Probabilistic Neural Network(PNN). They used different predictability based analysis techniques.

Takashi Kimoto et al[8] discussed a buying and selling timing prediction system for stocks on Tokyo Stock Exchange based on Modular neural network. They developed a number of learning algorithms and prediction methods for the TOPIX (Tokyo Stock Exchange Prices Indexes) prediction system. The prediction system achieved accurate predictions results .

III. MACHINE LEARNING

Machine learning is programming computers to optimize a performance criterion using example data or past experience. It has a model defined to some parameters, and learning is the execution of computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make prediction in the future, or descriptive to gain knowledge from data, or both.[9].Machine learning technique is alluring for artificial intelligence because it is based on principle of learning from training and experience. Machine learning has already gained an important place in trading and finance. There are many techniques that are used to predict movements in stock prices based on different models of time-series data of stock prices.

IV. ARTIFICIAL NEURAL NETWORK

A neural network is massively parallel distributed parallel processor made up of simple units(neurons) that has a natural propensity for storing experimental knowledge and make it available for use. It resembles the brain in two respects: Knowledge, acquired by the network from the environment through a learning process and inter-neuron connection strength, known as synaptic weight, used to store the acquired knowledge.[10]

Connection models such as ANN are well suited for machine learning where connection weights are adjusted to improve the performance of a network. Artificial Neural Network (ANN) is an information processing system where the elements called neurons, process the information. The signals are transmitted by means of connection links. The links possess an associated weight, which is multiplied along with the incoming signal (net input) for any typical neural network. The output signal is obtained by applying activations to the net input. The network consists of a set of sensory units that constitute the input layer and one or more hidden layer of computation modes. The input signal passes through the network in the forward direction. This type of network is called as multilayer Perceptron (MLP). The multilayer Perceptron are used with supervised learning.
and have to lead the successful back propagation algorithm where logistic sigmoid function is widely used. The MLP network has hidden neurons and this will make the network more active for complex tasks. The layers of network are connected by synaptic weights and have a high computational efficiency.

V. DATA AND METHODOLOGY

The purposed model has used Multilayer Feed-forward Neural Network and such types of networks consist of input layer, one or more hidden layers and an output layer. This research uses one input layer, one hidden layer and one output layer for stock price prediction. The model has been generated in five steps:

A. Data Collection

The actual problem is to forecast the stock price of Nifty 50 in India. For this purpose the dataset that has been used includes daily stock data of the National Stock Exchange which is available at National Stock Exchange’s Website www.nseindia.com. This research is based on secondary data. The data fields used in this research includes Open price, High price, Low price, Close price, Volume traded and Turnover of the Nifty 50 Index. The dataset taken encompasses the trading days from 03 April 2006 to 16 may 2016 and contains a total of 3335 instances. MBP version 2.2.4 has been used for this study. MBP is a free and open-source software which can be freely downloaded from http://dit.ipg.pt/MBP.

B. Data Pre-processing

The data must be prepared such that it covers the range of inputs for which the network is going to be used. Since the performance and reliability of the output from the neural network mainly depends on the quality of the data, therefore, the data must be pre-processed before it is fed to a neural network. First of all, we apply attribute relevance analysis on data so as to remove unwanted attributes from data and then the data will be normalized in the range -1 to 1 using min-max normalization technique. Since the input is in the normalized form, the output we get is also in the normalized form and hence, it must be de-normalized so as to have actual value. In order to train the network, dataset has been divided into two subsets:

1) Training Data Set(70%) : This data set has been used to train the network. The gradient was computed and biases and the weights of the connections between the neurons were adjusted accordingly.

2) Testing Data Set(30%) : This data has been used to test the performance of the network.

D. Training the Network

The network has been trained using back-propagation algorithm with the aim to improve the network performance i.e. to reduce root mean square error (RMSE). In this algorithm, the network has been trained by repeatedly processing the training data set and comparing the network output with the actual output and reducing the error to the minimum possible. If the error between network output and the actual falls below the threshold value, then the training stops otherwise weights of the connections between various neurons are modified so as to reduce “RMSE”. The modifications are done in the opposite direction i.e. from output layer through each hidden layer down to the first hidden layer. Since the modifications in the weights of the connections are done in the backward direction so the name given is back-propagation. Transfer functions calculate layer’s output from its net input. Hyperbolic tangent sigmoid transfer function and Log-sigmoid transfer function can be used for hidden layer and output layer.

E. Testing the Network

In this step the network has been tested using the test data set.
VI. EXPERIMENTAL RESULTS AND OBSERVATIONS

Table-1 depicts the RMSE values obtained for the proposed network model. The total training was for only 7001 epochs and the screenshots in Fig. 4 provide the detailed data about the training and testing patterns.

TABLE1. RMSE Errors

<table>
<thead>
<tr>
<th>S.no</th>
<th>Phase</th>
<th>Main Network</th>
<th>Space Network</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Training</td>
<td>0.0078485851</td>
<td>0.0041594316</td>
</tr>
<tr>
<td>2</td>
<td>Testing</td>
<td>0.0106902067</td>
<td>0.0055223479</td>
</tr>
</tbody>
</table>

Fig. 3, Fig.4, Fig.5 and Fig.6 correspond to network output for training and testing data after the training phase which shows that both the lines in graphs are almost superimposed upon each other. The results of network output were close to the desired values. The black lines represent the network output and the red lines represent desired output. It infers that the topology used in the network is optimal for the specified data.

Fig. 7 shows the weight and bias distribution from input layer to hidden layer and hidden layer to output layer.
The sensitivity analysis is the process which determines whether an input variable influences the output of the neural network or not. Fig. 8 shows the input sensitivity for main network. From the Figure it can be inferred that the 4th input which is closing price (C) influences output layer the most.

VII. CONCLUSION

Financier generally favors the stock market investments due to the opportunity of highest return. For companies, stock market is one of the main sources to gain money through initial public offer (IPO). This allows businesses to be run by public or raise extra capital for expansion by trading shares of ownership of the company in a public market. Indian stock market mainly consists of two major stock indices, Bombay Stock Exchange (BSE) and National Stock Exchange (NSE). The benchmark for these two exchanges are Sensex and Nifty. Nifty 50 is Index of NSE in India which encompasses top 50 listed companies of 23 different Sectors. Index is used to read the market sentiment by investors, traders and speculators. In this study, authors examined and applied multilayer Perceptron model by using the multiple back-propagation (MBP version 2.2.4) software for nifty 50 index forecasting. The results from analysis shows that by using MBP software Nifty 50 index can be predicted with reasonable accuracy and offer the ability to predict the stock prices more accurately. By using this tool one can have the ability to forecast the stock price of NSE Nifty 50 more accurately. This analysis can be used to reduce the RMSE (Root mean square error) in predicting the future stock prices. It increases the chances for the investors to predict the prices more accurately by reducing the error percentage and hence increase their profit in share markets. The model has high prediction accuracy and fast
convergence speed. The predicted results are satisfactory and can be used for effective decision making.

REFERENCES


AUTHOR’S BIOGRAPHY

Gourav Kumar is currently pursuing Master of Technology in Computer Science from Department of Computer Science & IT, University of Jammu, Jammu, India. His research interest includes Artificial Intelligence, Machine Learning, Neural Network, and Data Mining.

Dr. Vinod Sharma is professor of Computer Science and currently holds the position of Head, Department of Computer Science and IT, University Of Jammu, J&K, India. He holds a PhD in Computer Science and has more than 25 publications in various journals and conferences. His research areas of interest are Artificial Intelligence (AI), Machine Learning, Databases and Data Mining.