ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

Stock Market Index Prediction by Hybrid Neuro-Genetic Data Mining Technique

Ganesh V. Kumbhar[1], Rajesh V. Argiddi[2]

Research Scholar, Computer Science & Engineering Department, WIT, Sholapur, India[1]

Assistant Professor, Computer Science & Engineering Department, WIT, Sholapur, India[2]

Abstract: Stock Exchange Share Price is very hard to predict because of fluctuating nature of stock. Stock price prediction is one of the emerging fields of research and many methods like technical analysis, statistical analysis, time series analysis etc are used for this purpose. The aim of this study is to develop a system that can predict future stock prices based on samples of historical prices. This study will use the Artificial Neural Network for learning the historical data of Stock Market & make predictions for next few days. Genetic algorithm is used to optimize the weights of Neural Network so that it can learn better. We are using the stock historical data of 10 years from 2004 to 2014. This data is used to build the system & we use the system for predicting the stock behaviour for next few days. The result showed that the proposed model gives better results than simple Neural Network model.

Keywords: Data Mining, Genetic Algorithm, Neural Network, Stock Price

I. INTRODUCTION

With the increased economic globalization and rapid evolution of information technology, the financial data are being generated and accumulated at an unusual pace. As a result, for effective and efficient utilization of massive amount of financial data, there has been a need for automated approaches to support industries and individuals in strategic planning and decision-making. Data mining techniques are used to discover hidden patterns and predict future trends in financial markets. The competitive advantages of use of data mining techniques include increased revenue and much improved decision making capability.

When studying financial markets the main issue is of course predicting price movements, but these are very noisy systems, being influenced by economical and political factors such as companies' earnings, political issues, news or natural disasters. As such they are extremely hard to predict, nonetheless investors widely use market analysis techniques to study and forecast market movements. These methods are technical analysis which studies the price and volume of the assets, using past information to predict the future behaviour, and fundamental analysis that deals with various economic and politic factors, looking down from the global economy all the way to the company itself.

Time-series forecasting is a special type of classification on which we concentrate. Specifically, for any financial time series related to the performance of an individual stock, the goal is to predict the value of the time series *k* steps into the future.

Stock market prediction is the act of trying to determine the future value of a company stock or other financial instrument traded on a financial exchange. In the previous work the different techniques were proposed like SVM, ARIMA, GA, NN, Association Rule mining & different machine learning technique for prediction of stock market. Among these techniques we are combining the features of GA & NN for stock prediction. Thus we propose a method to predict the stock price using a hybrid genetic approach combined with NNs. We describe a number of input variables that help the network to train & forecast the next day price; a GA is used for optimizing NN's weights.

We present a new system that uses genetic algorithms (GAs) & Neural Network to predict the future behaviour of individual stocks. More generally, the system uses GAs for optimization of weights of neural network while training. Then the trained Neural Network is used to predict the future behaviour of stock market

II. GENETIC ALGORITHM

Genetic Algorithms (GA) are direct, parallel method for global search and optimization. GA is part of the group of Evolutionary Algorithms (EA). The Genetic Algorithms are direct, stochastic method for optimization. Since they use populations with allowed solutions (individuals), they count in the group of parallel algorithms.

After an initial population is randomly generated, the algor ithm evolves the through three operators:

1) selection which equates to survival of the fittest
2) crossover which represents mating between individual
3) mutation which introduces random modifications

SELECTION OPERATOR

Selection process selects better individuals & passes on their genes to the next generation.

The selection of each individual depends on its fitness. Fitness is determined by function which may be objective function or a subjective judgement.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

CROSSOVER OPERATOR
By using selection operation two individuals are chosen from the population. Then randomly chooses a crossover site along the bit strings. The values of the two strings are exchanged up to this point.

If S1=11110000 and s2=00001111 and the crossover point is 2 then S1'=00110000 and s2'=11001111.

The two new offspring are created from this mating and those are put into next generation of the population. By recombining portions of selected good individuals, this process creates even better individuals.

MUTATION OPERATOR
Some of bits of a portion of the new individuals will have flipped with some low probability. It maintains diversity within the population and inhibits premature convergence. Mutation alone induces a random walk through the search space.

Basic block diagram of Genetic Algorithm is

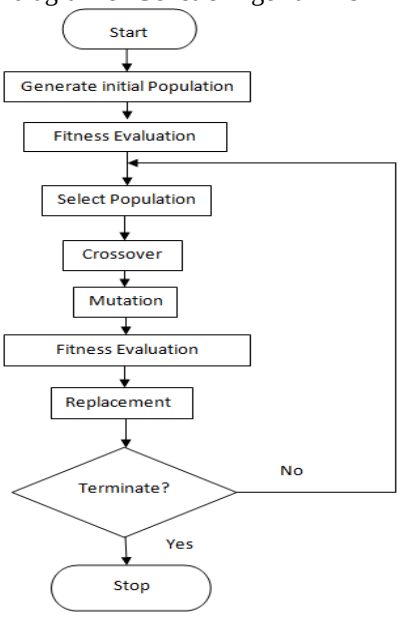

Fig. Genetic algorithm

Genetic Algorithm is a randomized algorithm that could be run for a very long time to obtain an optimal solution.

Neural Network

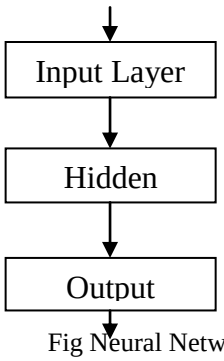

Fig Neural Network

The basic architecture of ANN is Multi- layer perceptron . In this architecture, information flows mainly in one direction (from input to output). It consists of one input, one or more hidden and one output layer. Inputs are sent into units in input layer then weighted output from these units is taken as input in next hidden layer, weighted output of this layer is sent as input in next hidden layer and so on until output of last hidden layer is sent to output layer. Output layer gives the predicted output.

III. LITERATURE REVIEW

In the previous research, different techniques were presented for historical stock data analysis & stock market price prediction.

Yung-Keun Kwon and Byung-Ro Moon [1] proposed a new technique to predict the stock price using a hybrid genetic approach combined with recurrent NNs. For that they described a number of input variables that help the network to forecast the next day price. Technical indicators or signals that were developed in deterministic trading techniques are used for input variables. The back propagation algorithm is prone to get stuck in local minima and highly depends on the initial weights; a GA is used for optimizing NN's weights.

Kai Keng Ang, Chai Quek [2] used neuro-fuzzy systems and neural networks for forecasting stock price difference on artificially generated price series data. In that they proposed a novel rough set-based neuro-fuzzy stock trading decision model called stock trading using rough set-based pseudo outer-product (RSPOP) which synergize the price difference forecast method with a forecast bottleneck free trading decision model. This proposed stock trading forecast model uses the pseudo outer-product based fuzzy neural network using the compositional rule of inference with fuzzy rules identified using the RSPOP algorithm as the underlying predictor model and simple moving average trading rules in the stock trading decision model.

Jorge Fonseca, Rui Neves, Nuno Horta [5] proposed an investment strategy using Genetic Algorithms applied to the stock market. In order to build a portfolio of promising stocks they have done fundamental analysis by using indicators such as earnings volatility and growth, Price-to-Earnings ratio and Price/Earnings to Growth ratio. In addition with this technical indicators such as moving average crossovers and Relative Strength Index are used to adapt the portfolio to the market's trends.

Ramin Rajabioun and Ashkan Rahimi-Kian [3] proposed a genetic programming model in which ,first a precise mathematical model is obtained for four competing or cooperating companies stock prices and then the optimal buy/sell signals are ascertained for five different agents which are trading in a virtual market and are trying to maximize their wealth over one trading year period. The model is so that gives a prediction of the next 30th day stock prices.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

Yusuf Perwej, Asif Perwej [9] proposed a method to predict the daily excess returns of Bombay Stock Exchange (BSE) indices over the respective Treasury bill rate returns. They have applied the prediction models of Autoregressive feed forward Artificial Neural Networks (ANN) to predict the excess return time series. For the Artificial Neural Networks model using a Genetic Algorithm is constructed to choose the optimal solution.

Kyoung-jae Kim, Ingoo Han [10] proposed a genetic algorithms (GAs) approach to feature discretization and the determination of connection weights for artificial neural networks (ANNs) to predict the stock price index. GA not only searches for the optimal or near-optimal solutions of connection weights in the learning algorithm but also looks for the optimal or near optimal thresholds of feature discretization for the dimensionality reduction. GAFD discretizes the original continuous data according to the GA-derived thresholds and simultaneously assigns the genetically evolved connection weights.

Hyun-jung Kim , Kyung-shik Shin [11] proposed the effectiveness of a hybrid approach based on the adaptive time delay neural networks (ATNNs) and the time delay neural networks (TDNNs) with the genetic algorithms (GAs) in detecting temporal patterns for stock market prediction tasks. By evaluating the fitness of different sets of the number of time delays and network architectural factors at the same time, they proposed a solution for the ATNN and TDNN model.

O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L.J. Herrera, A. Guillen, L. Marquez, M. Pasadas [12] proposed a hybrid ARIMA–ANN model for time series prediction. This model combines the ARIMA models, and the power of ANNs.

In literature review we have reviewed different techniques for stock market prediction such as ARIMA, GA, NN, Association Rule mining & different machine learning technique for prediction of stock market.

IV. METHODOLOGY

Our proposed methodology is a two stage model. In first stage we create artificial neural network & learn it with stock market historical data .The Genetic algorithm is used to optimize the weights of neural network so that it can learn the pattern of stock market very efficiently. In the second stage we use the learned neural network to predict the future nature of stock market that is whether it goes up ,down or it remains steady for next few days.

The below diagram shows the first stage of proposed methodology where we create the Artificial NN which is recurrent NN. The weights of NN are sent to Genetic algorithm as an input which initializes its population with the weights. Then GA does the selection, crossover & mutation on population & returns the best population with weights of NN. The optimized weights are set to NN & it is tested for the desired output .If desired output is found then it stops Genetic Algorithm otherwise GA continues until termination condition is met. Once we get solution we save the NN & Use it for future prediction which is second stage of proposed methodology.

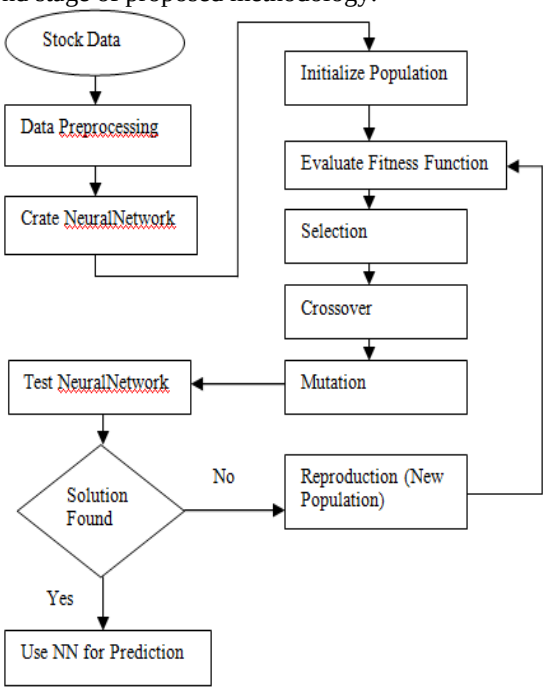

Fig. System Architecture

Model Analysis:

There are two main phase

1) Training phase
2) Prediction phase or testing phase

The first phase training phase is further divided into two parts, building ANN model and weight updating phase.

Here we are using recurrent neural network which consists of 3 layers that are one Input layer, one hidden layer, one output layer. The structure of neural network used is (5,9,1) which means the input layer consist the 5 input nodes , hidden layer consist 9 hidden nodes & output layer consist 1 output node. The input parameters provided to the neural network are

1. previous day's close price
2. open price
3. high price
4. low price
5. close price

The output of neural network is the predicted Closing Stock Price. Further a program is written, which predicts whether stock price goes up or down by comparing the predicted values with real values.

We have used only one hidden layer and after trying out 5-15 numbers of nodes we have found the number of nodes in the hidden layer. The best result we have found for the number of nodes in the hidden layer equals to 9.

We have used sigmoid function as activation function.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

WEIGHT UPDATION USING GENETIC ALGORITHM:
Here we have used Genetic algorithm Optimization to update the weight of our Neural Network model.

Genetic Algorithm trains the neural networks using the minimum mean squared error as stopping criterion for learning, while never exceeding the maximum number of generations which can take testing data and predicts the future stock closing values.

DATA SET USED:
Historical Stock Price data (from 01.01.04 to 31.12.14) of Google (GOOG), TCS(TCS.BE), INFOSYS(INFY.BO), HCLTECH (HCLTECH.NS),WIPRO(WIPRO.BO)collected from http://in.finance.yahoo.com

RESULT:
Here we have used 70% of historical data for training of the NN model & remaining 30% for testing. The NN model constructed predicts the behaviour of stock price that is whether it goes up or down. Also we have compared the constructed Hybrid GANN model with simple NN model & result showed that GANN model works better than simple NN model.
Following table gives the result that we have got for Google.

GOOG	NN	GANN	Actual
Date	Behaviour	Behaviour	Behaviour
30-12-2014	Down	Down	Down
29-12-2014	Up	Down	Up
26-12-2014	Up	Up	Up
24-12-2014	Down	Down	Down
23-12-2014	Up	Up	Up
22-12-2014	Up	Up	Up
19-12-2014	Up	Up	Up
18-12-2014	Up	Down	Down
17-12-2014	Down	Up	Down
16-12-2014	Down	Down	Down
15-12-2014	Down	Down	Down
12-12-2014	Up	Up	Up
11-12-2014	Up	Down	Down
10-12-2014	Down	Down	Down
09-12-2014	Down	Down	Down
08-12-2014	Up	Down	Down
05-12-2014	Down	Down	Down
04-12-2014	Up	Down	Up
03-12-2014	Up	Up	Up
02-12-2014	Up	Up	Up
01-12-2014	Up	Down	Up
28-11-2014	Up	Up	Up
26-11-2014	Down	Down	Down
25-11-2014	Down	Down	Down
24-11-2014	Up	Up	Up
21-11-2014	Up	Up	Up
20-11-2014	Up	Up	Up
19-11-2014	Up	Up	Up
18-11-2014	Up	Up	Up
17-11-2014	Up	Up	Down
14-11-2014	Up	Up	Up
13-11-2014	Down	Up	Down
12-11-2014	Up	Down	Down
11-11-2014	Down	Down	Down
10-11-2014	Up	Up	Up
07-11-2014	Down	Down	Down
06-11-2014	Up	Up	Up
05-11-2014	Up	Down	Up
04-11-2014	Up	Up	Up
03-11-2014	Down	Down	Down
31-10-2014	Up	Up	Up
30-10-2014	Down	Down	Down
29-10-2014	Up	Down	Down
28-10-2014	Down	Up	Up
27-10-2014	Down	Down	Down
24-10-2014	Up	Up	Up
23-10-2014	Up	Up	Up
22-10-2014	Up	Up	Up
21-10-2014	Down	Down	Down
20-10-2014	Down	Up	Down
17-10-2014	Down	Down	Down
16-10-2014	Down	Down	Down
15-10-2014	Down	Down	Down
14-10-2014	Up	Down	Up
13-10-2014	Down	Down	Down
10-10-2014	Down	Up	Up
09-10-2014	Up	Down	Down
08-10-2014	Up	Up	Up
Prediction	83.050847	84.482759	
Accuracy:			

Similarly we have also got the results for TCS, INFOSYS, HCLTECH and WIPRO. The result showed the hybrid GANN model works better than simple NN model.

V. CONCLUSION

As researchers and investors strive to find the behaviour of market, the use of Neural network algorithm to forecast stock market prices will be a continuing area of research. We have proposed the hybrid neuro-genetic technique to predict stock market price. Genetic algorithm is used for optimization of weights of neural network & NN is used for prediction. Here we have used previous day's close price, open price, high price, low price, close price as inputs of ANN. If instead of these inputs, different economical measures like Moving average, Moving average convergence and divergence ,General Index, Net Asset Value, Relative strength index ,P/E ratio, Earning per Share, Share Volume will be used as inputs to ANN, then our approach of weight updation using Genetic algorithm may give better result.

DOI 10.17148/IJARCCE.2015.4571

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 5, May 2015

In conclusion we can say that if we train our system well efficiently with historical data set it will generate more accurate & error free prediction.

REFERENCES

[1]. Yung-Keun Kwon and Byung-Ro Moon., "A Hybrid Neurogenetic Approach for Stock Forecating." IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 18, NO. 3, MAY 2007.
[2]. Kai Keng Ang, Chai Quek, "Stock Trading Using RSPOP: A Novel Rough Set-Based Neuro-Fuzzy Approach" , IEEE TRANSACTIONS ON NEURAL NETWORKS, 2006.
[3]. Ramin Rajabioun and Ashkan Rahimi-Kian, "A Genetic Programming Based Stock Price Predictor together with Mean-Variance Based Sell/Buy Actions", Proceedings of the World Congress on Engineering 2008 Vol II, WCE 2008, July 2 - 4, 2008, London, U.K.
[4]. Osman Hegazy , Omar S. Soliman and Mustafa Abdul Salam, "A Machine Learning Model for Stock Market Prediction" , International Journal of Computer Science and Telecommunications ,Volume 4, Issue 12, December 2013
[5]. Jorge Fonseca, Rui Neves, Nuno Horta , "Optimizing Investment Strategies based on Companies Earnings using Genetic Algorithms", published by ACM GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands
[6]. Preeti Paranjape-Voditel , Dr.Umesh Deshpande ," An Association Rule Mining based Stock Market Recommender system". 2011 Second International Conference on Emerging Applications of Information Technology, Published by the IEEE Computer Society.
[7]. Michael Hagenau, Matthias Hauser, Michael Liebmann, Dirk Neumann, "Reading all the news at the same time: Predicting mid-term stock price developments based on news momentum" , 2013 46th Hawaii International Conference on System Sciences ,Published by the IEEE Computer Society
[8]. Zabir Haider Khan, Tasnim Sharmin Alin, Md. Akter Hussain, "Price Prediction of Share Market using Artificial Neural Network(ANN)". International Journal of Computer Applications (0975 – 8887) Volume 22– No.2, May 2011
[9]. Yusuf Perwej, Asif Perwej , "Prediction of the Bombay Stock Exchange (BSE) Market Returns Using Artificial Neural Network and Genetic Algorithm", Journal of Intelligent Learning Systems and Applications, 2012, 4, 108-119
[10]. Kyoung-jae Kim, Ingoo Han. "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index", Expert Systems with Applications 19 (2000) 125–132
[11]. Hyun-jung Kim , Kyung-shik Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets" , Applied Soft Computing 7 (2007) 569–576
[12]. O. Valenzuela, I. Rojas, F. Rojas, H. Pomares, L.J. Herrera, A. Guillen, L. Marquez, M. Pasadas. "Hybridization of intelligent techniques and ARIMA models for time series prediction" . Fuzzy Sets and Systems 159 (2008) 821 – 845.