Forecasting the Tehran Stock Market by Artificial Neural Network

(CasestudyMobarakeh-steelCo.)

Reza Aghababaeyan, TamannaSiddiqui, NajeebAhmadKhan
Department of Computer Science
Jamia Hamdard University
New Delhi- India

Abstract— One of the most important problems in modern finance is finding efficient ways to summarize and visualize the stock market data to give individuals or institutions useful information about the market behavior for investment decisions. The enormous amount of valuable data generated by the stock market has attracted researchers to explore this problem domain using different methodologies. Potential significant benefits of solving these problems motivated extensive research for years. In this paper, computational data mining methodology was used to predict seven major stock market indexes. Two learning algorithms including Linear Regression and Neural Network Standard feed-forward back prop (FFB) were tested and compared. The models were trained from four years of historical data from March 2007 to February 2011 in order to predict the major stock prices indexes in the Iran (Tehran Stock Exchange). The performance of these prediction models was evaluated using two widely used statistical metrics. We can show that using Neural Network Standard feed-forward back prop (FFB) algorithm resulted in better prediction accuracy. In addition, traditional knowledge shows that a longer training period with more training data could help to build a more accurate prediction model. However, as the stock market in Iran has been highly fluctuating in the past two years, this paper shows that data collected from a closer and shorter period could help to reduce the prediction error for such highly speculated fast changing environment.

Keywords- Data mining; *Stock Exchange; Artificial Neural Network; Matlab.*

I. INTRODUCTION

Data mining called knowledge discovery in databases, in computer science, the process of discovering interesting and useful patterns and relationships in large volumes of data. The field combines tools from statistics and artificial intelligence (such as neural networks and machine learning) with database management to analyze large digital collections, known as data sets. An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspect of biological neural networks. Neural Networks (NN) as Artificial Intelligence method has become very important in making stock market predictions, as it has proved to be more advantages than the other methods. Since

then lot of research was carried out using different topologies of Neural Networks. According to Wong, Bodnovich and Selvi [1] the most frequent areas of neural networks applications are production operations (53.5%) and finance (25.4%). In finance, it is more specific to stock market predictions. When compared to the other methods NN outperformed and the accuracy rate ranges from 68% to 90% [9]. The popularity of these methods is mainly due to the benefits outnumber the limitations it has. Different NN methods were used for optimal feature selection to generating buy and sell signals, the more popular being the former.

In this paper, we anticipated Mobarakeh-Steel Co. try. The high level was in Tehran Stock Exchange. We used data from 15 March 2007 until 14 February 2011 for training the neural Network and from15 February 2011 until 30th we have performed experiments using MATLAB and we got 97 % results, which are very encouraging.

In stock market when brokers want to sell or buy stock, they mostly depend on technical trading rules. Robert Edward and John Magee have [10], defined technical trading rules as "the science of recording the actual history of trading (price changes, volume of transaction, etc.) in a certain stock or in 'The averages' and then deducing from that pictured history the probable future trend". Different artificial intelligence methods were used to optimize the prediction by successful selection of trading rules.

II. PROPOSED APPROACH

A researcher originates primary data for the specific purpose of the problem at hand, but secondary data are data that have already been collected for other purposes. Secondary data can be classified as internal and external. Our research strategy is the analyses are of secondary data. For conducting our research, types of data are needed:

1) Companies' stock prices, which are an internal secondary data, gathered from the financial databases.

2) 2. Existing Approaches to Cash Forecasting.

Techniques used for cash forecasting can be broadly classified.

a) Time -series Method

b) Factor analysis method

c) Expert system approach

We have used Factor analysis method for cash forecasting because by using this method the results are better than of two above methods.

Among the 20 text files provided by Tehran Stock Exchange Service Company (TSESC), Mobarakeh-Steel Co. Intra day price and their corresponding date and time during years 2007 till 2011 are chosen to be given to the split and merge algorithm. Before implementing, the segmentation algorithm in R Programming Language, Mobarakeh-Steel Co. text file should be read by the program. The program reads 1904 intra day prices (data points) for this company during years 2007 and 2011, which is equal to 2266.

TABLE 1.   Data Sets

Training Set	From 15nd march 2007 to14th February 2011 for training the neural
Validation Set	From 15 February 2011 to 30th February 2011for validating the neural

Architecture of the Model:

Design of right architecture involves several important steps:

a) Selecting the number of layers

b) Basic decision about the amount of neurons to be used in each layer

c) Choosing the appropriate neurons' transfer functions.

If there are not enough neurons in each layer, the outputs will not be able to fit all the data points (under- fitting). On the other hand, if there are too many neurons in each layer, oscillations may occur between data points (over-fitting). Therefore, a topology study was conducted in order to find the most appropriate architecture neural network to fit Cash forecasting parameters. There are several combinations of neurons and layers.

III.   EXPERIMENTAL

The Propose approach has been implemented through Matlab Software.

A. Number of Layers

The model is a three layer feed-forward neural network and was trained using fast back propagation algorithm because it was found to be the most efficient and reliable means to be used for this study. Table 2 shows a comparison of the two algorithms.

B. Input Layer Size

Number of input neurons depends upon:

1) Number of cash withdrawal factors included in the model

2) Way these factors are encoded

TABLE 2.   Selection of Algorithms

Function	Technique	Time(s)
TRAINBP	Back propagation	59
TRAINBPX	Fast Back propagation	40

Mainly calendar effects are included as parameters affecting cash withdrawal in this model, Total number of input neurons needed in this model hence is seven, each representing the values of an individual variable at a particular instant of time.

C. Output Layer Size

In this model, only one output unit is needed for indicating the value of forecasted cash.

D. Optimal Hidden Layer Size

There is no easy way to determine the optimal number of hidden units without training using number of hidden units and estimating the error of each. The best approach to find the optimal number of hidden units is trial and error. In practice, we can use either the forward selection or backward selection to determine the hidden layer size.

Forward selection: Starts with choosing an appropriate criterion for evaluating the performance of the network. Then we select a small number of hidden neurons; record its performance i.e. forecast accuracy.

Next, we slightly increase the hidden neurons, train and test until the error is acceptably small or no significant improvement is noted.

Backward selection: Starts with a large number of hidden neurons and the decreases the number gradually.

For this study, the forward selection approach was used to select the size of hidden layer and best result was with 10 Neurons in hidden layer, as evident from the following Table 3.

TABLE 3.   Optimal Neurons in Hidden Layer

No. of Neurons in Hidden Layer	Mean Forecast Error for 60datapoints	Sum Squared error at 5000epochs
7	9.675	0.004343
8	9.85726	0.004153
9	8.51395	0.002881
10	9.52748	0.003768
11	9.39397	0.003565
12	8.56169	0.003645

Optimal Transfer Function As shown in table 4, for best transfer function, tansig in hidden (at 10 neurons) and logsig in output layer were found to be optimal.

TABLE 4.   Optimal Transfer Function

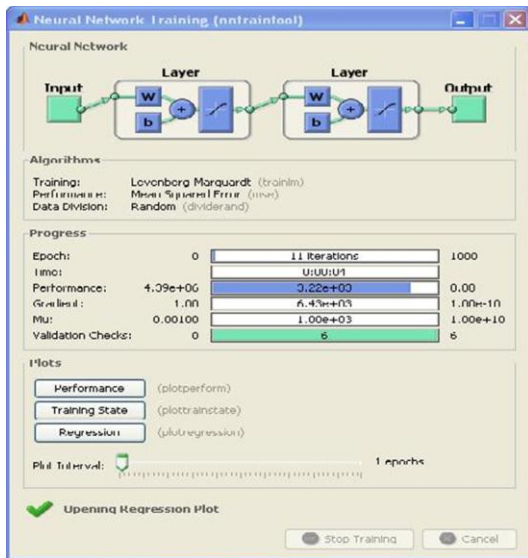| Transfer function	Mean Forecast Error for 60datapoints	Sum Squared error at 5000epochs
Logsig - purelin	11.0786	0.007697
Tansig - purelin	9.90157	0.004647
Logsig - logsig	7.38031	0.002591
Tansig - tansig	9.53892	0.004560
Logsig - tansig	12.2373	0.011609
Tansig - logsig	6.74556	0.002246

Figure1, Neural Network Training' Shown in the figure 1 Mobarakeh-steel co. data training, that data training was 15nd march 2007 to 14th February 2011 for Training.
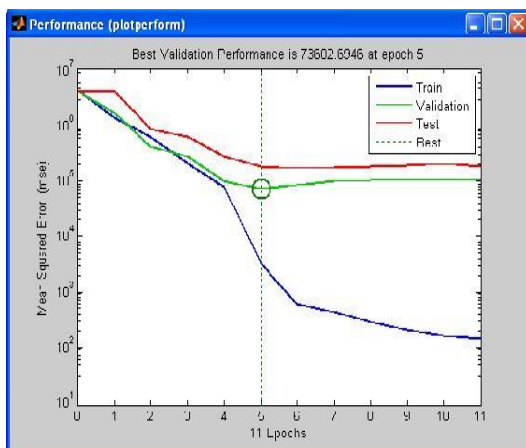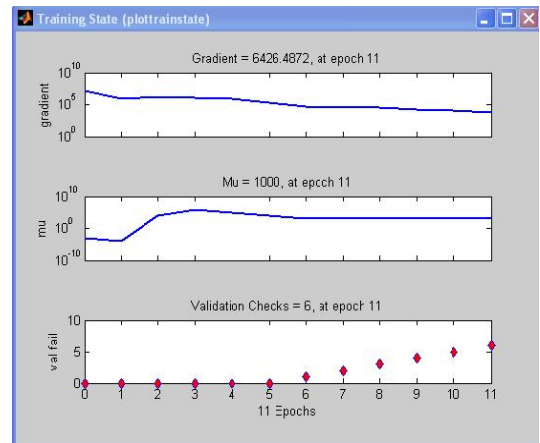

Figure2, Performance

Figure3, Training State 'Shown in the figure 3 Training State (Plot train state), that Gradient= 6426.4872 at epoch 11. Mu = 1000, at epoch 11. Validation checks= six, at epoch 11.
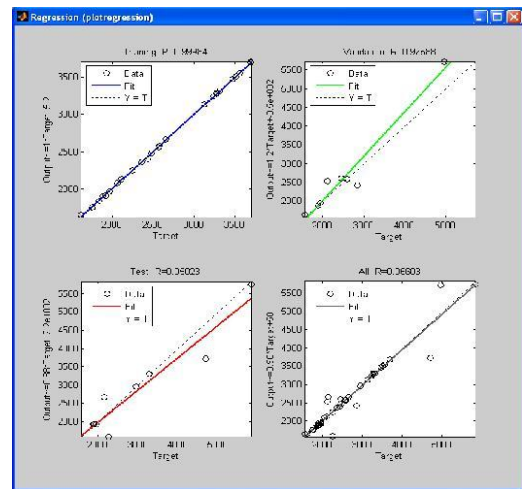

Figure4, regression (plot regression) 'Shown in the figure 4 Regression= 0.96603

TABLE 5. Real Price of mobarakeh-steel co

1904	1890	1845	2061	2435	2949	3128	3557
3279	3273	3470	3275	3310	3514	4696	5819
4957	3689	3689	1889	1965	1926	1754	1766
1882	2122	1560	1920	1920	2118	1948	1626
1895	1967	2253	2848	2462	2367	2153	2591
2652	2580	2475	2564	2564	3241	2266	

Shown in the figure above Real Price of Mobarakeh-steel co. from 15nd march 2007 to 14 February 2011.

In this model, only one output unit is needed for indicating the value of forecasted cash. Shown in the table 6 that percentage predicted for Real Data. Shown in the table 7 that Prices Predicted from Mobarakeh-steel Co.
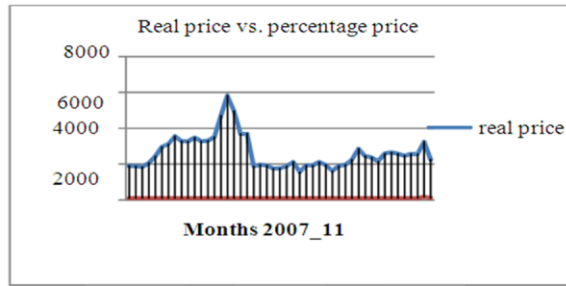

Figure 5, Real price vs. percentage price.

TABLE 6.   Percentage of Prediction

1826	1823	1815	2063	2354	2894	3087	3493
3241	3235	3419	3243	3275	3446	4663	5349
4947	3634	3634	1912	1939	1925	1789	1791
1815	2114	1697	1905	1905	2092	1928	1723
1885	1945	2240	2863	2453	2323	2082	2573
2649	2560	2431	2540	2540	5497	2713	

TABLE 7. Price Prediction of mobarakeh-steel Co.

96	97	98	100	97	98	99	98
99	99	99	94	99	98	99	92
100	98	99	101	99	100	102	102
97	100	109	99	99	99	99	106
100	99	100	101	100	98	97	99
100	99	98	99	99	169	117	

IV. RESULTS AND DISCUSSIONS

Results: In this work, we anticipated Mobarakeh-Steel Company's real price data. The high level was in Tehran Stock Exchange. We used data from 15 March 2007 till 14th February 2011 for training the neural Network and from 15 February 2011 until 30 February 2011 for validating the network. After the network was created by MATLAB, the results were 97 percent, which were very encouraging for this research work.

V. CONCLUSION

An Overview of Study:

Stock markets have been studied over and over again to extract useful patterns and predict their movements. Mining textual documents and time series concurrently, such as predicting the movements of stock prices based on the contents of the intraday price, is an emerging topic in data mining and text mining community. Stock price trend forecasting based solely on the technical and fundamental data analysis enjoys great popularity. However, numeric time series data only contain the event and not the cause why it happened.

Textual data such as news articles have richer information, hence exploiting textual information especially in addition to numeric time series data increases the quality of the input and improved predictions are expected from this kind of input rather than only numerical data. Information about company's report or breaking news stories can dramatically affect the share price of a security.

In order to make the prediction model, the research process should be implemented consists of different steps including data collection, data preprocessing, alignment, feature and document selection, document representation, classification and model evaluation. With the prediction model (feed-forward back prop) we can conclude that our prediction model out performs the random labeling. The prediction model will notify the up or down of the stock price movement when an upcoming pieces of news is released, and 83 percent of time can predict correctly.

This can be very beneficial for individual and corporate investors, financial analysts, and users of financial news. With such a model, they can foresee the future behavior and movement of stock prices; take correct actions immediately and act properly in their trading to gain more profit and prevent loss.

REFERENCES

[1] B.K. Wong, T.A. Bonovich, Y. Selvi, "Neural Network Applications in Business: A Review and Analysis to the literature (1988-95), Decision Support Systems, Vol. 19, Pp301-320, 1997.

[2] C. Moler, the creator of MATLAB (December 2004). "The Origins of MATLAB". Retrieved April 15, 2007.

[3] C. Moler, "MATLAB Programming Language". Retrieved 2010-12-17.

[4] C. Moler in a Math works newsletter Cleve Moler, the creator of MATLAB (2000). "MATLAB Incorporates LAPACK". Retrieved December 20, 2008.

[5] D.R. Cooper, P.S. Schindler, 2003. Business Research Methods. 8thed .New York: Mc Graw-Hill.

[6] D. Spiel man, (2004-02-10). "Connecting C and Matlab". Yale University, Computer Science Department. Retrieved 2008-05-20.

[7] J.Kazama, T.Makino, J. Tsujii, "Support Vector Networks. Machin Learning" (1995), 20(3), p.273-297

[8] G. Amos (2004). MATLAB: An Introduction with Applications 2nd Edition. John Wiley & Sons. ISBN 978-0-471-69420-5.

[9] M. Zelaic, "Neural Network Applications in Stock Market Prediction –

A Methodology Analysis" Proc. of 9th Intl'Conf. Information and Intelligent Systems, 1998.

[10] R. Edward, J. Magee, 'Technical Analysis of Stock trends'. Seventh Edition 1997.

[12] R. Goering, "Matlab edges closer to electronic design automation world," EE Times, 10/04/2004

[13] Wikipedia,"MATLAB technical documentation". Mathworks.com. Retrieved 2010-06-07.

[14] J. Stafford, "The Wrong Choice: Locked in by license restrictions,"SearchOpenSource.com, 21 May 2003.

AUTHORS PROFILE

Reza Aghababaeyan, PhD Student of Computer Science, New Delhi-India 2011.

Tamanna Siddiqui, Ph.d of Computer Science New Delhi-India (my supervisor).

Najeeb Ahmad Khan, Master of Computer Science New Delhi-India.