ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Forecasting short-term stock prices using sentiment analysis and Artificial Neural Networks

Binoy B Nair[1]*, Kumar PN[2], Sidharth R Prasad[1], Lopa M Singh[1], Vijayalakshmi K[1], Sai Ganesh R[1], Reshma J[1]

[1]Dept. of Electronics and Communications Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India.

[2]Dept. of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, Amrita University, India

***Corresponding author E-mail: b_binoy@cb.amrita.edu**

ABSTRACT

Short-term fluctuations in stock prices are generally considered to be extremely difficult to predict, primarily due to their nonlinear nature. The authors believe that one of the reasons for such seemingly unpredictable fluctuations is the type of sentiment prevailing amongst traders at that point in time. An attempt has been made in this study to forecast the stock returns using the sentiments expressed on social media and Artificial Neural Networks. The proposed system is validated on stocks drawn from the Indian stock markets. Results indicate that the proposed technique can indeed be successfully used for short-term forecasting of stock prices.

KEY WORDS: Artificial Neural Networks, Sentiment, Forecasting, Stock.

1. INTRODUCTION

Forecasting short-term fluctuations in stock prices is considered to be a challenge by both traders and researchers alike, mainly due to the fact that such fluctuations tend to be nonlinear and hence, difficult to predict. The Efficient Market Hypothesis, as proposed in (Fama, 1970) suggests that market efficiency makes it impossible to forecast stock prices. Earlier studies on the subject, e.g. (Cowels, 1933; 1944), also cast doubts on the possibility of successfully forecasting stock prices. More recent studies, e.g. (Atsalakis, 2009; Binoy, 2015), however, have demonstrated that it is indeed possible to successfully forecast stock prices. It is also observed from (Binoy, 2011; 2015; 2010), that soft computing based techniques tend to generate better results when compared to the traditional technical analysis or statistical techniques. However, there have been very few studies on the impact of popular sentiments on the fluctuations in stock prices. Social media has been used for real world event content detection (Hila, 2011; Takeshi, 2010), identification of current news topics of interest (Jeongin, 2013; Alan, 2012), and new event detection (Petrovic, 2010). Attempts have also been made to predict stock prices based on interpretation of financial news articles (Robert, 2006). However, it was observed that there are very few studies on the impact of prevailing sentiments on the stock prices. It is also observed that study of social media sentiments on Indian stocks has not been reported so far, hence, the study is unique in this regard.

In the present study, a system that can categorize the sentiments about a particular stock into good, bad or neutral, and forecast the returns based on the overall sentiments using Artificial Neural Networks (ANNs) is presented. The proposed system is validated on stocks drawn from the Indian stock markets. Rest of the paper is organized as follows: Section 2 presents the design of the proposed system; results are presented in Section 3 and conclusions in Section 4.

2. PROPOSED SYSTEM

The social media platform considered for the study is Twitter. Twitter is an online social media service where users can post and read messages referred to as tweets. There were three major reasons for choosing Twitter in the present study: a) The tweets can be sent and received almost on a real-time basis, b) The length of each tweet is limited to a maximum of 140 characters, making the preprocessing and filtering of tweet data less resource-intensive, c) There are around 302 million monthly active users with 500 million tweets being sent per day, as of May 2015, thus there is no dearth of sentiment data.

With so many people tweeting about various opinions on subjects ranging from cosmetics to multi-national companies, twitter is a source of ample real time information about the current trends and news.
Block diagram of the proposed system is presented in Fig.1.

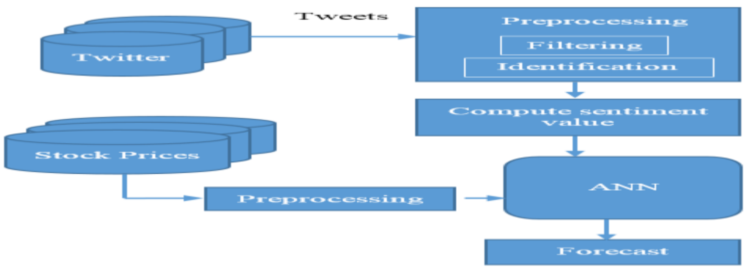

Fig.1.Proposed System Block Diagram

Steps in the design of the proposed system are as follows:

Preprocessing: The first step in the preprocessing phase is tweet extraction and filtering. From amongst the millions of tweets, only those that contain references to the specific firms whose stock prices are to be forecast, are filtered and extracted. Large-cap stocks randomly drawn from the Indian stock markets are considered in the present study.

The stocks considered are: Delhi Land & Finance Ld. (DLF), ITC Ltd. (ITC), Nestle, Oil and Natural gas Corporation (ONGC), National Thermal Power Corporation (NTPC), Maruti-Suzuki India Ltd. (MSIL), Tata Consultancy Services (TCS), State Bank of India (SBI), Reliance Industries Ltd. (RIL), Housing Development Finance Corporation Ltd. (HDFC) and Airtel.

It was decided to choose only large-cap firms for the present study since stocks of these firms tend to be highly liquid (resulting in very frequent and rapid price fluctuations) and are widely discussed (resulting in availability of higher number of tweets for the study). Next step is the identification of tweets relevant to the forecasting process. This accomplished by checking for the presence of certain specific keywords in the tweets. In the present study, the selected tweets are grouped into three categories: Good, Bad and Neutral.

The process of identification of the tweet type for each day is presented in the algorithm find_tweet_type below:

```
Algorithm find_tweet_type
        Begin
                i,j ← 1
                Good, Bad, Neutral ← 0
While i < total_tweets do
    While j < length (tweet(i) ) do
        If tweet (i, j:j+2)= "buy"
                Else If tweet (i, j:j+9)= "accumulate"
                        Then    Good ← Good+1
                End
        End
        If tweet (i, j:j+3)= "sell"
                Else If tweet (i, j:j+5)= "reduce"
                        Then    Bad ← Bad+1
                End
        End
        If tweet (i, j:j+3)= "hold"
                        Then    Neutral ← Neutral+1
                End
     End
End
        End Algorithm
```

It must be noted that in the above algorithm, issues such as matching of the exact case of words in the tweet and the possibility of array indices going out of bounds are not considered, however, they must be incorporated in any program that is written based on the above algorithm. The variable total_tweets represents the total number tweets selected after the filtering process and the length () function calculates the total number of characters in the tweet.

Computing Sentiment Value: Once the tweets are segregated, the overall sentiment value about a particular stock on a particular day t is identified as follows:

$$\mathrm{S(t) = Good - Bad} \qquad (1)$$

Where Good and Bad are calculated from the algorithm find tweet type.

Stock Price Preprocessing: The stock price preprocessing involves the computation of the daily return. The return value on the day t is given by R(t) and is computed as follows:

$$R(t) = \frac{y_t - y_{t-1}}{y_{t-1}} \qquad (2)$$

Where, y_t = Closing price of t^{th} day, y_{t-1} =Closing price of $t\text{-}1^{th}$ day

ANN based forecasting of returns: In the present study, a single hidden layer feed forward ANN trained using the Levenberg-Marquardt (LM) training algorithm is employed for one-day ahead forecasting of stock returns. It has been observed that LM algorithm is best suited for training the ANN and hence, has been used in the present study as well. The architecture of the network used in the study is presented in Fig. 2.

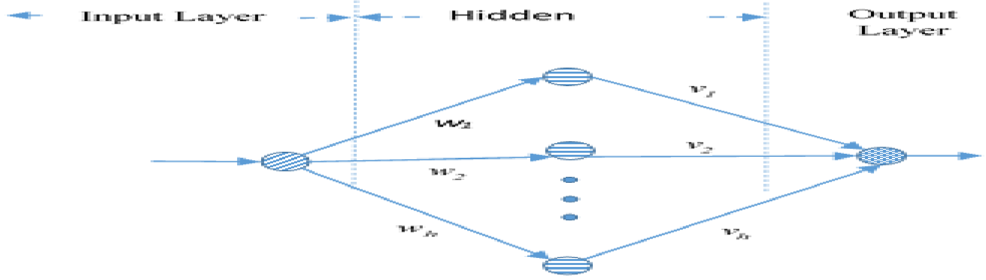

Fig.2. ANN architecture employed

In Fig.2, the j-th input sample is represented by I_j and the corresponding output by O_j. There is only one hidden layer considered with the optimal number of hidden layer neurons (h) being selected by trial and error. In the Fig.2, the input neuron is connected to each hidden layer neuron by a weight w_{1i} , where $1 \le i \le h$ and each hidden layer neuron and the output neuron has bias weights $b_1, b_2, \ldots, b_h$ and b_0 respectively. The input to hidden layer neuron i, denoted by I_{hi} for the input I_j is then given by:

$$I_{hi} = w_i \, I_j + b_i \qquad (3)$$

The output of the hidden neuron i, denoted by O_{hi} is given by:

$$Out_{hi} = f(I_{hi}) \qquad (4)$$

Where f (.) is the activation function.

The output to the output neuron O_j is then given by:

$$O_j = f(\, b_0 + \textstyle\sum_{i=1}^{h} v_i . Out_{hi}) \qquad (5)$$

The ANN training parameters used are presented in Section 3.

3. RESULTS AND ANALYSIS

The proposed system was validated on eleven stocks drawn from the Indian stock markets, as listed above. The details of the tweets and the time frames are presented in Table 1. It must be noted that in table 1, the tweets are collected from 27/3/2015-16/4/2015.

Table.1.Tweet details

Firm	No. of tweets selected	Firm	No. of tweets selected
Airtel	172	Nestle	161
DLF	180	NTPC	100
HDFC	91	ONGC	106
SBI	461	RIL	47
MSIL	30	ITC	32
TCS	67		

The ANN parameters considered in the present study were:

Number of inputs: 1, Number of outputs: 1, Number of hidden neurons: selected by trial and error, Activation function for the output and hidden layer neurons: tan-sigmoid, Where $tan\ sigmoid\ (n) = \frac{2}{1+e^{-2n}} - 1$, Activation function for the input neuron: linear, Learning algorithm: LM, Minimum gradient for termination: 10^{-7}, Maximum number of epochs: 1000, Performance goal: 0, Minimum μ : 0.001, Maximum μ : 10^{10}, μ increase factor: 10, μ decrease factor: 0.1.

The whole dataset is divided into training: testing = 90: 10. The optimal number of hidden neurons was obtained using trial-and-error such that the Pearson's correlation coefficient between the actual and predicted returns (R) is maximized. The performance of the proposed system for the stocks considered are presented in Table 2.

Table.2.Proposed system performance

Firm	Optimal Hidden Neurons	R	Firm	Optimal Hidden Neurons	R
Airtel	80	0.998323	RIL	60	0.862996
DLF	40	1	ITC	80	0.56002
HDFC	80	0.999674	NTPC	60	0.811368
Nestle	40	1	MSIL	80	0.900308
TCS	40	0.976039	ONGC	80	0.919496
SBI	60	0.443901			

4. CONCLUSIONS

It is observed from the results that the number of hidden neurons needed for optimal performance varies from stock to stock. It is also observed that the proposed system is able to generate high forecasting accuracy in nine out of the eleven stocks considered. It is believed that the proposed system will help the investors in making an informed decision about their trade.

REFERENCES

Alan Ritter, Oren Etzioni, and Sam Clark, Open domain event extraction from twitter, 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012, 1104-1112.

Atsalakis G.S, and Valavanis K.P, Surveying stock market forecasting techniques – part II: Soft computing methods, Expert Systems with Applications, 36 (3), 2009, 5932–5941.

Binoy B Nair, S Gnana Sai, A N Naveen, A Lakshmi, G S Venkatesh and V P Mohandas, A GA-Artificial Neural Network Hybrid System for Financial Time Series Forecasting, in Lecture Notes in Computer Science-Communications in Computer and Information Science, 147 (3), 2011, 499-506.

Binoy B Nair, V P Mohandas, Nikhil Nayanar, E S R Teja, S Vigneshwari, and K V N S Teja, A Stock Trading Recommender System Based on Temporal Association Rule Mining, SAGE Open, 5 (2), 2015, 1 –10.

Binoy B Nair and Mohandas V P, An Intelligent Recommender System for Stock Trading, Intelligent Decision Technologies, 9, 2015, 243–269.

Binoy B Nair, and Mohandas V.P, Artificial Intelligence Applications in Financial Forecasting- A Survey and Some Empirical Results, Intelligent Decision Technologies, 9, 2015, 99-140.

Binoy B Nair, Mohandas V.P, and Sakthivel N.R, A Genetic Algorithm Optimized Decision Tree-SVM based Stock Market Trend Prediction System, International Journal on Computer Science and Engineering, 2 (9), 2010, 2981-2988.

Binoy B Nair, Mohandas V.P, and Sakthivel N.R, Predicting stock market trends using hybrid ant-colony-based data mining algorithms: an empirical validation on the Bombay stock exchange, International Journal of Business Intelligence and Data Mining, 6 (4), 2011, 362-381.

Binoy B Nair, Patturajan M, Mohandas V.P, and Sreenivasan R.R, Predicting the BSE sensex: performance comparison of adaptive linear element, feed forward and time delay neural networks, International Conference on Power, Signals, Controls and Computation, Trissur, 2012, 1-5.

Cowles A, Can stock market forecasters forecast?, Econometrica, 1 (3), 1933, 309-324.

Cowles A, Stock market forecasting, Econometrica, 12, 1944, 206-214.

Fama E.F, Efficient capital markets: A review of theory and empirical work, Journal of Finance, 25 (2), 1970, 383-417.

Hila Becker, Mor Naaman, and Luis Gravano, Beyond Trending Topics: Real-World Event Identification on Twitter, ICWSM 11, 2011, 438-441.

Jeongin Kim, Byeongkyu Ko, Huijin Jeong, and Pankoo Kim, A method for extracting topics in news twitter, 7th International Journal of Software Engineering and Its Applications, 7 (2), 2013, 1-6.

Johan Bollen, Huina Mao, Zeng, and Xiaojun, Twitter mood predicts the stock market, Journal of Computational Science, 2 (1), 2011, 1-8.

Petrovic S, Osborne M, and Lavrenko V, Streaming first story detection with application to Twitter, in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010, 181-189.

Robert Schumaker, and Hsinchun Chen, Textual analysis of stock market prediction using financial news articles, Twelfth Americas Conference on Information Systems, Acapulco, Mexico, 2006, 1432-1440.

Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, Earthquake shakes Twitter users: real-time event detection by social sensors, 19th international conference on World Wide Web, 2010, 851-860.

Twitter Inc. (2015, May) Twitter Website. [Online]. https://about.twitter.com/company