A Detail Survey on Predicting Stock Price Movement based on Communication Network

Shital Shiral
Department of Computer Science & Engineering
VVPIET, Solapur University, Solapur, India

Harshal Torvi
Department of Computer Science & Engineering
VVPIET, Solapur University, Solapur, India

ABSTRACT

Stock price prediction is a popular topic in financial studies. Stock market is basically nonlinear in nature and predicting share price is very difficult because there are no specific set of rules to estimate the price of the share in share market. Many methods are used to predict the share price like statistical analysis, time series analysis but none of these methods are considered to be consistently acceptable prediction methods and applying traditional methods may not ensure the accuracy of prediction. Various machine learning algorithms have been used to study the highly unpredictable nature of stock market by capturing repetitive patterns. Various companies have their preferred analysis tool for stock market forecasting and the reason for preference is the accuracy with which they predict. This paper gives brief survey of well-known prediction techniques used for prediction of stock in the stock market.

General Terms

Stock market pattern recognition, Stock market prediction algorithm

Keywords

Share Market, Artificial Neural Network, Time series

1. INTRODUCTION

Stock Market prediction is an interesting topic for researchers because there is no specific method that has been designed to accurately predict stock price movement despite of numerous scientific attempts that have been made. Stock Market prediction helps to determine the future value of a company stock and several other financial instruments traded in a stock exchange. Stock market plays a vital role in the economy of country [4]. Two main entities are involved in the stock market first is investors and second is industry and both the entities are interested to know whether some stock will rise or fall over certain period of time. For any company to raise funds for business expansion, stock market is a primary source. Higher the demand for company's stock then higher will be the share price of company and share price of company will be low if demand of company's stock is low [6].

There are two approaches to carry out prediction, fundamental approach and technical approach [12]. In fundamental analysis method uses numeric information like profit ratios and earnings for prediction while technical analysis utilize modeling techniques such as graphs to predict trends in price. Even with the lack of prediction methodologies there is little success in prediction of stock price, both methodologies are rely on historical data to predict the stock market price [8][9]. Textual data can also help Stock Market prediction. Data for prediction can be extracted from company's news stories and quarterly reports. Stock market prediction using text mining is heavily depending on predefined set of keywords and some machine learning rules [17]. In this method weights are

assigned to keywords with respect to the movement of a share price. Some text mining prediction tools retrieve financial news articles and blogs published in previous year and respective stock market prices for same period. Retrieved articles are then labeled as positive or negative depending on their impact on stock price movement. Classification tools like support vector machines is then trained with labeled train articles. This train model is then used to predict the future stock price of the company [11] [16].
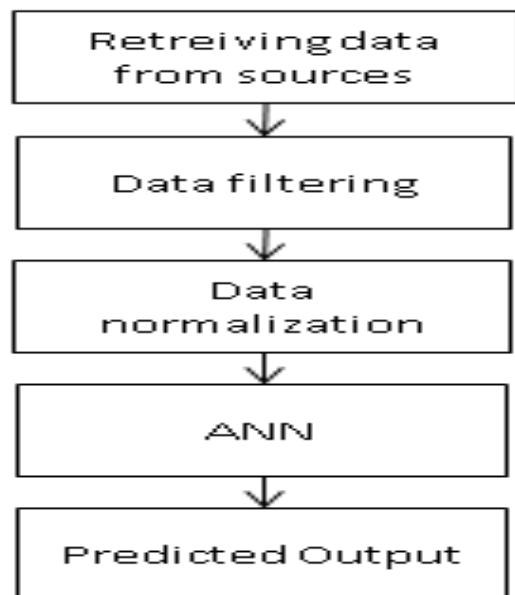

Fig 1 Prediction Process

There is a lot of considerable text mining work done on prediction of stock prices. Work basically contains text categorization models targeting the prediction of stock price movement. Fig. 1.1 shows the traditional prediction process using neural network. There are several classification methods are used in stock prediction such as k nearest neighbor, naive Bayes and support vector machines. Support vector machine is most widely used method for classification but most of the classification methods have accuracy rate below 60 % because of nature of stock price movements and the fact that it is immensely chaotic.

Section I contains the introduction of stock price movement prediction and its practices, Section II contains a survey on the related work and existing frameworks, Section III contains detail description of proposed system and architecture, Section IV concludes research work with future directions).

2. LITERATURE SURVEY

[1] This paper states that a corporate e-mail ecosystem contains meaningful information about employees' communication network and patterns. Author has only focused on the communication frequency that is e-mail exchanged between employees. Author used Enron dataset in which he identified patterns of e-mail exchange. Author believes that such patterns can reveal important information about organizations stability that may subsequently influence the company's performance in the stock market.

Author used this methodology to indirectly predict company's stock performance. Finally author stated that there is a correlation between email communication network and stock price in Enron dataset and can be accurately predicted based on the detected relationships.

Advantages:

1. Simple system architecture

2. Prediction rule can not only provide condition part and conclusion part, but also present the weighted part for the rules.

3. E-mail communication data and stock price data is treated as multidimensional sequential or time series data. This proposed algorithm can be used to handle similar kinds of time-sequential data.

Limitations:

1. Only used communication frequency between different people in the e-mail communication network.

2. Possibility of false positive rate if email context is negative

[2] In this paper author proposed an intelligent time series prediction system. To construct a stock price forecasting expert system author used meta heuristic firefly algorithm and least squares support vector regression (MetaFA-LSSVR). System uses sliding-window meta heuristic optimization and is a graphical user interface that runs as a stand-alone application. As compared to other prediction methodologies proposed methodology is very simple and involves very few computations. To generalize the application of proposed system author used stock price market in Taiwan and can be extended to analyze multivariate time series data and estimate other stock markets such as China, Korea, and Vietnam etc.

Advantages:

1. Involve fewer computations

2. Work best for highly non-linear time series, whose patterns are difficult to capture by traditional models

Limitations:

1. Computational speed is very low with respect to sliding-window validation, because of the complexity of solving large mathematical loops.

[3] To build prediction module author proposed LSTM (Term Memory Long-Short) recurrent neural networks. To analyze the stock data author has first filtered the data extracted feature value. The main work of LSTM recurrent neural networks is to extract feature value and analyze the stock data. Author stated that the proposed model can play a better forecasting effect, even though the accuracy is not very high

as per the experimental results. Results are calculated on dataset of JoinQuant platform. System achieved 72% of accuracy for the short period of data. Author believes that the model has a lot of space to improve its accuracy and if the more features to train the model are provided then system can achieve desirable accuracy.

Advantages:

1. Model has better forecasting effect

2. Computation process is faster

Limitations:

1. Prediction accuracy is not very high

[4] In this paper before developing the model author first studied various methods to find the most accurate model for prediction of prices of the stock. To get better results author increased nodes of neural network as stated by author of paper "Stock Price Prediction and Trend Prediction Using Neural Networks". Author found that Feed Forward Neural network provides the highest accuracy for the stock prediction. Finally author calculated error of results in which least amount of mean absolute percentage error is 1.81598342% for feed forward neural network and the maximum error is 11.32847594% which is obtained using linear model with polynomial trend.

The result obtained was the opening price of the stock and that too was average for a full month. So an improvement in this system can be achieved by forecasting the opening price of each day.

Advantages:

1. System implements Feed Forward Network which provides highest accuracy for stock prediction

2. System also perform well on raw data without filtering

Limitations:

1. Gives only long term predictions

[5] In this paper author aims to prove that artificial neural network is an effective tool to predict market stock. Paper provides comparative study between Long Short-Term Memory (LSTM) model and Multilayer Perceptron (MLP) model. Proposed heuristic method of analysis and study followed in this paper are optimal for stocks which are highly frequently traded and can outperform (EMH) Efficient Market Hypothesis theory. 10 unique stocks recorded on New York Stock Exchange are used to calculate results. Calculated results prove that MLP has outperformed LSTM model, in predicting short term stock prices.

Advantages:

1. Accuracy is high on short term intervals

Limitations:

1. System requires trading at extremely short time intervals nullifying time required for decision making so prediction accuracy for long time intervals is low

[6] In this paper author proposed a stock recommendation system using J48 and random forest algorithm. Author firstly optimized the stock price prediction only for short term, author found that using specific pre-processing tools and machine learning framework it is possible to improve accuracy of short term stock prediction but these technical

indicators are not that helpful to predict long term stock trend but some other factors like historical data that is net profit of a company of previous years, PE ratio and promoters holding can help to predict long term stock price.

Advantages:

1. Model gives prediction with considerable accuracy for short term as well as long intervals

Limitations:

1. Raw data needs to be sampled before processing for prediction.

[7] In this paper author studied the usage of Long Short Term Memory (LSTM) network to predict future trends of stock prices. System uses technical analysis indicators along with previous stock price data. Author used real data from Brazilian stock exchange. Author then evaluated series of experiments and the results were analyzed against various machine learning algorithms and stock market prediction tools. Results depicted that system is predicting the price of the stock with an accuracy of 55.9%.

Advantages:

1. Work best for non-linear time series data with an accuracy of 55.9%

Limitations:

1. System carried out complex mathematical loops so the computational speed is low.

[8] In this work, author studied various stock market prediction algorithms before building a pr3ediction model. Proposed prediction model gives monthly prediction as well as daily prediction ro forecast the next day stock market price. Author used Sentiment analysis approach to identify and extract sentiments from individual user in social media platform. Correlation between sentiment and stock price is determined. For classification of sentiments and related stock value is done by using three algorithms separately Support Vector Machine, Artificial Neural Network and Multiple Linear Regression and comparative analysis is done on the result generated by all three algorithms.

Advantages:

1. Author used ANN. Artificial Neural Network is the classification algorithm for deep learning. The result shows that the deep learning algorithm performs better than the MLR and SVM

Limitations:

1. Computational complexity is an issue.

[10] This paper proposes an application of ARIMA model Author stated that ARIMA model help to predict the future stock indices which have a great influence on the performance of the Indian economy. Author mainly focused on Indian stock market because author believes that Indian Stock market is the center of interest for many researchers and economists as well as investors. To develop the model author applied the validation technique with the observed data of Sensex of 2013.

Advantages:

1. Bet fitted for liner forecasting

Limitations:

1. In case of change in the Government policies like monetary or expert input policies will result in higher fluctuation in Sensex. In that case, prediction of Sensex using proposed model will not give desirable result.

[11] In this paper author considered both technical and fundamental analysis approaches. Machine learning algorithm is applied on historical stock prices in technical analysis and in fundamental analysis sentiment analysis is applied on social media data Author believes that social media data has perform a key role while predicting the trend of the stock market. In fundamental analysis method involves collecting social media data, financial news blogs and extracting sentiments expressed by users. Then the correlation map is calculated between sentiments and stock values

Advantages:

1. System has big data analytical capabilities.

Limitations:

1. Sentiment analysis uses summative assessment of the sentiments in a news blogs or tweet; this could be improved for better sentiment calculations.

3. PROPOSED METHODOLOGY

Using the social communication network within company among employees, the proposed algorithm can analyze the relationship between communication context and the movements (high and low) of stock price. We assume that the values of movements of stock price are related to the values of frequency of communication in the previous time point.
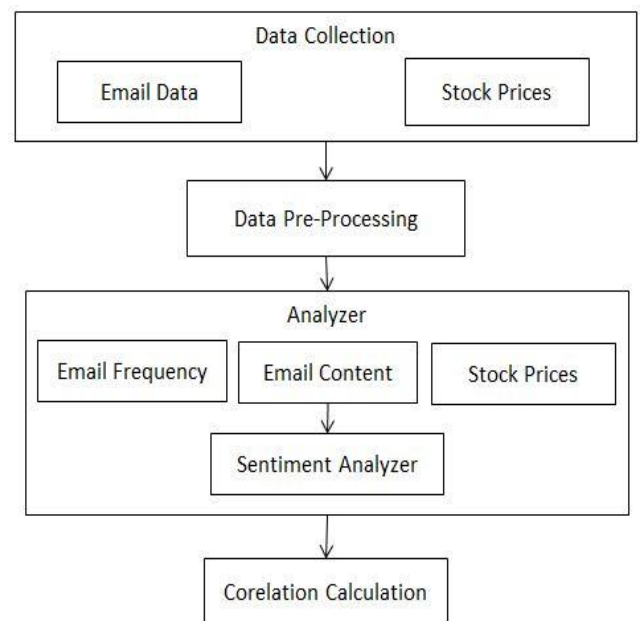
Fig. 2 System design for proposed system

The task of our proposed algorithm is to uncover these temporal relationships and predict the movements of stock price using communication network. The proposed algorithm includes the following steps:

1. Discretizing the weight matrices and the value of movement

2. Discovering patterns for describing the relationship between communication frequency and stock price

3. Constructing predication rules based on patterns

4. Predicting the movement of stock price using prediction rules

The purposed method for developing the system consists of mainly three main steps. Firstly, data is collected and sorted for relevancy. Secondly, analysis is carried out on the collected data by examining the email flow, email content after which the data is represented and scored accordingly. At last, an ANN is designed and a suitable algorithm yielding best accuracy is chosen to predict the stock value

3.1 Modules Description

3.1.1 Data Collection
In this module data is collected as per the time point from dataset. Here we are using Enron dataset. When Enron collapsed in 2001, about 500.000 internal emails were made public. We are using cleaned Enron data and that is in MSQL format.
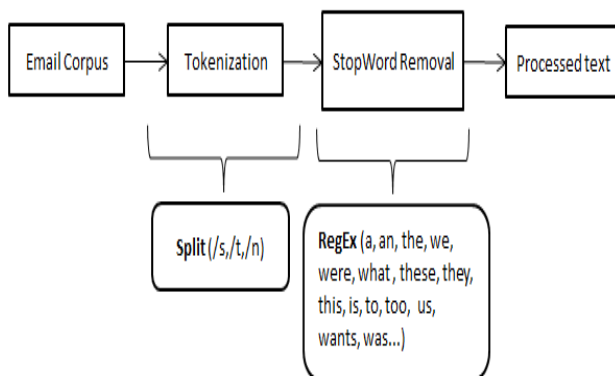
3.1.2 Data Pre-Processing

Fig. 3 Data Pre-Processing

3.1.2.1 Tokenization:
Email body is split into individual words based on the space and irrelevant symbols are removed. We form a list of individual words for each email body.

In fig. 3.2 we are using tokenization, we are using Split function with some parameters. Parameters that we are providing are "/s","/t","/n"

Table 1 Regular expression

Regular Expression	Description
\s	stands for "whitespace character"
\n	Refers **to** "new line" (aka one tab vertical space)
\t	Refers **to** "four horizontal spaces"

3.1.2.2 Stop word Removal:
Words that do not express any emotion are called stop words. Words like a, is, the, with etc. are removed from the list of words. To remove stop words we are using Regex function. We are providing dictionary of stop wards to Regex function to match those words and remove them from statements.

3.1.3 Analyzer

3.1.3.1 Sentiment Analysis:
Emails are classified as positive, negative and neutral based on the sentiment present. 70% email body out of the total emails are examining manually and annotated as 1 for Positive, 0 for Neutral and 2 for Negative emotions. For classification of nonhuman annotated emails a machine learning model is trained whose features are extracted from the human annotated emails. The extracted features are fed to the classifier and trained using LibSVM algorithm.

3.1.4 communication frequency calculation
This module calculates the communication frequency that is mail exchange count between two nodes within given time interval.

The proposed algorithm includes the following steps:

1. Discretizing the weight matrices and the value of movement;

2. Discovering patterns for describing the relationship between communication frequency and stock price;

3. Constructing predication rules based on patterns;

4. Predicting the movement of stock price using prediction rules

3.1.5 Correlation calculations
This is a final stage of proposed system. All features extracted from analyzer module (sentiment result value, communication frequency value) is then provided to LibSVM. When the model with LibSVM is trained with 90 percent of data, it is expected to give a result of 80%. These results give a significant edge to the investors and they show good correlation between stock market movements and the communication between company networks.

4. CONCLUSION
Stock exchange prediction helps 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.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Pei-Yuan Zhou , Keith C.C. Chan, Member, IEEE, and Carol Xiaojuan Ou, "Corporate Communication Network and Stock Price Movements:

[2] Insights From Data Mining", IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS, 2018.

[3] Jui-Sheng Chou and Thi-Kha Nguyen "Forward Forecast of Stock Price Using Sliding-window Metaheuristic-optimized Machine Learning Regression". IEEE 2018

[4] Siyuan Liu1 , Guangzhong Liao1 ,Yifan Ding1Hubei Key Laboratory of Intelligent Information Processing and Real-time Industrial System Stock Transaction Prediction Modeling and Analysis Based on LSTM IEEE 2018.

[5] Shashank Tiwari Akshay Bharadwaj Dr. Sudha Gupta "Stock Price Prediction Using Data Analytics. 978-1-5386-3852-1/17/$31.00 2017 IEEE"

[6] Kaustubh Khare, Omkar Darekar, Prafull Gupta, Dr. V. Z. Attar Department of Computer En-gineering and Information Technology, College of Engineering, Pune."Short Term Stock Price Prediction Using Deep Learning IEEE 2017"

[7] Mr.Rupesh A. Kamble Computer science and engineering Government College of Engineering Aurangabad, India Kamblerupesh9@gmail.com "Short and Long Term Stock Trend Prediction using Decision Tree" ICICCS IEEE 2017.

[8] David M. Q. Nelson, Adriano C. M. Pereira, Renato A. de Oliveira "Stock Market's Price Movement Prediction With LSTM Neural Networks" IEEE 2017.

[9] R. Yamini Nivetha, Dr. C. Dhaya Department of Computer Science and Engineering. Adhiparasakthi Engineering College Melmaruvathur, Kanchipuram District "Developing a Prediction Model for Stock Analysis" IEEE 2017.

[10] ASHISH SHARMA, DINESH BHURIYA, UPENDRA SINGH "Survey of Stock Market Prediction Using Machine Learning Approach" IEEE 2017.

[11] Debadrita Banerjee Student, Department of Statistics St.Xavier's College Kolkata, India deba.gb@gmail.com "Forecasting of Indian Stock Market using Time-series ARIMA Model" IEEE 2014

[12] Girija V Attigeri, Manohara Pai M M, Radhika M Pai, Aparna Nayak Manipal Institute of Technology, Manipal-576104, India "Stock Market Prediction: A Big Data Approach"IEEE 2014

[13] Aparna Anant Bhat, Sowmya Kamath S. Department of Information TechnologyNational Institute of Technology, KarnatakaSurathkal, India "Automated Stock Price Prediction and TradingFramework for Nifty Intraday Trading" IEEE 2013

[14] Radu IacominFaculty of Automatic Control and ComputersUniversity POLITEHNICA of BucharestBucharest, RomaniaEmail: raduiacomin@yahoo.com "Stock Market Prediction"IEEE 2015

[15] Tao Xing, Yuan Sun, Qian Wang, Guo Yu International School of Software Wuhan University Wuhan, China e-mail:xingtao@whu.edu.cn, cjchsun@hotmail.com "The Analysis and Prediction of Stock Price"IEEE 2013

[16] Zhao, LeiBaylor UniversityEmail: cxhdy@foxmail.comWang, LinJapan Advanced Institute of Science and TechnologyEmail: linwang@jaist.ac.jp "Price Trend Prediction of Stock Market UsingOutlier Data Mining Algorithm" IEEE 2015

[17] Vishal S. Shirsat,Rajkumar S. Jagdale,S. N. Deshmukh, Department of Computer Science and IT Dr. B. A. M. University, Aurangabad, India "Document Level Sentiment Analysis from News Articles" 978-1-5386-4008-1/17/$31.00 ©2017 IEEE

[18] Xiaobo Zhang, Qingsong Yu School of Computer Science and Software Engineering East China Normal University Shanghai, China e-mail: Bobe_24@126.com, qsyu@cc.ecnu.edu.cn "Hotel Reviews Sentiment Analysis Based on Word Vector Clustering" 2017 2nd IEEE International Conference on Computational Intelligence and Applications

[19] Snehal Kale, Vijaya Padmadas ,Department of Computer Science & Engg. Thodomal Shahani Engineering College Mumbai, India "Sentiment Analysis of Tweets Using Semantic Analysis" 978-1-5386-4008-1/17/$31.00 ©2017 IEEE

[20] M.Trupthi, Suresh Pabboju, G.Narasimha, Computer Science Department, JNTUH, Jagital, Telangana State, India "SENTIMENT ANALYSIS ON TWITTER USING STREAMING API" 2017 IEEE 7th International Advance Computing Conference

[21] Ana Valdivia, M. Victoria Luzón, and Francisco Herrera, University of Granada "Sentiment Analysis in TripAdvisor" 1541-1672/17/$33.00 © 2017 IEEE

[22] Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain,Department of Computer Science and Engineering, Thapar University, Patiala "STATISTICAL AND SENTIMENT ANALYSISOF CONSUMER PRODUCT REVIEWS" IEEE 8th ICCCNT 2017

[23] Mauro Dragoni,Fondazione Bruno KesslerGiulio Petrucci,Fondazione Bruno Kessler, University of Trento E-mail: [dragoni,petrucci]@fbk.eu "A Neural Word Embeddings Approach ForMulti-Domain Sentiment Analysis" OI 10.1109/TAFFC.2017.2717879, IEEE Transactions on Affective Computing

[24] Chlóe Clavel and Zoraida Callejas "Sentiment analysis: from opinion mining tohuman-agent interaction" DOI 10.1109/TAFFC.2015.2444846, IEEE Transactions on Affective Computing

[25] Shenghua Liu, Xueqi Cheng, Fuxin Li, and Fangtao Li "TASC:Topic-Adaptive Sentiment Classificationon Dynamic Tweets" DOI10.1109/TKDE.2014.2382600, IEEE Transactions on Knowledge and Data Engineering

[26] Wei Zhao, Ziyu Guan∗, Long Chen, Xiaofei He,Fellow, IAPR,Deng Cai, Beidou Wang and Quan Wang "Weakly-supervised Deep Embedding for Product Review Sentiment Analysis" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. ?, NO. ?, AUGUST 2017

[27] Abdullah Alfarrarjeh 1, Sumeet Agrawal 2, Seon Ho Kim 3, Cyrus Shahabi 4Integrated Media Systems Center, University of Southern California, Los Angeles,

CA 90089, USA "Geo-spatial Multimedia Sentiment Analysis in Disasters" IEEE 2017 International Conference on Data Science and Advanced Analytics

[28] Nishant Suman, P K Gupta, Pankaj Sharma Department of Computer Science and Engineering Jaypee University of Information Technology, Solan, "Analysis of Stock Price Flow Based on Social Media Sentiments" IEEE 2017 International Conference on Next Generation Computing and Information Systems

[29] Rushlene kaur bakshi, Navneet Kaur, Ravneet kaur, Gurpreet Kaur "Opinion mining and sentiment analysis" 978-9-3805-4421-2/16/\$31.00c©2016 IEEE

[30] Aliza Sarlan1, Chayanit Nadam2, Shuib Basri3Computer Information Science University Teknologi PETRONAS, Perak, Malaysia,aliza_sarlan@petronas.com.my; hayanit171@gmail.com; shuib_basri@petronas.com. "Twitter Sentiment Analysis" IEEE 2014 International Conference on Information Technology and Multimedia