International Journal of Recent Engineering Research and Development (IJRERD)
Volume No. 01 – Issue No. 03, ISSN: 2455-8761
www.ijrerd.com, PP. 01-04

Application of Sentiment Analysis in Stock Markets

Abhishek Chander NV[1], CH.Vanipriya[2]
Sir M Visvesvaraya Institute of Technology, Bangalore, Karnataka

Abstract: Investment in stock is assumed to be the way of growing money quickly. But the problem is deciding upon which stock to invest in. Without any intelligence, investing in the stock would be gambling. But, thanks to researchers, it is widely believed that market sentiment, the sentiments that are hidden in the reviews or posts regarding that stock, are highly influential in determining the actual price. The main aim of this research is to provide a qualitative assessment of the correlation between historical stock prices and market sentiment at the time.

Keywords: Sentiment analysis, Stock markets

I. INTRODUCTION

Any investment in stock markets is done with an expectation of rapid appreciation of the investment. However the stock market is a wildly unpredictable system that resists most attempts at modeling it. This inherent uncertainty leads to apprehensive investors, who are unsure of the markets and their investment's future. Complete surety in their investment is not within any investor's reach; however, experience and intimate understanding of the markets certainly increase the odds of a successful return.

The investors attempt to chart the trajectory of a stock, ideally buying when it is at the lowest and selling it when it reaches its peak price. Traditional statistical models are used widely in economics for time series prediction. These statistical models are restricted to modeling linear relationships between factors influencing markets and the value of market. Historical pricing models involve analysis of historical stock data to identify patterns that are then extended to predict future stock prices. Overall Market sentiment is determined with a variety of methods such as number of growing vs declining stocks and new peaks versus new lows comparison. Market sentiment has been said to play a major role in the overall movement of an individual stock. Various prediction techniques used for stock price prediction are, traditional time series prediction, neural networks and State Vector Machines etc.

Earlier, the main source of stock related information was newspapers and experts were consulted for advice. But the trend has started to shift recently as internet use has proliferated at a rapid pace. As of November 2015, India had 375 million internet users, making it the second-largest Internet user-base in the world overtaking United States of America. The major Indian stock markets introduced Internet trading (online-trading) in February 2002. Investors have started to incorporate market sentiment into their buying decision. Market sentiments are measured using news analytics, which include analyzing sentiment of news articles related to companies and sectors.

With the widespread adoption of online trading, an overwhelming amount of data is available on the internet which can be consulted. The available data can be mainly classified into two, numerical data, like historical prices, and textual data, like news articles and reviews by online media. The majority of earlier work on stock prediction has been focused on analysis on the numerical data like application of time series prediction on historical prices.

In recent years the trend has been to consult various news items and expert advice online before investing. Textual data, like expert advice, have an inherent tone or emotion attached to it, a sentiment. An expert's views on the subject may be positive owing to its good performance and its potential. Evaluation of large volume of text manually is not only tedious and time consuming but may also yield inconsistent results. Sentiment analysis aims to automate this process.

II. MOTIVATION

The motivation behind this work is, an article in a news paper how the opinions and reviews expressed by the people on some products in the review sites, blogs and discussion forums are affecting the overall sales of their products. Another reason was a malicious attack which was held against ICICI Bank. In 2008, the stock value of ICICI was dipping suddenly and many customers were withdrawing their money and started closing their accounts in that bank. These were caused by rumors asking account holders of ICICI Bank to withdraw their deposits. It was noticed that the rumors were spreading via SMS and internet. These rumors led to a major loss of faith in the bank.

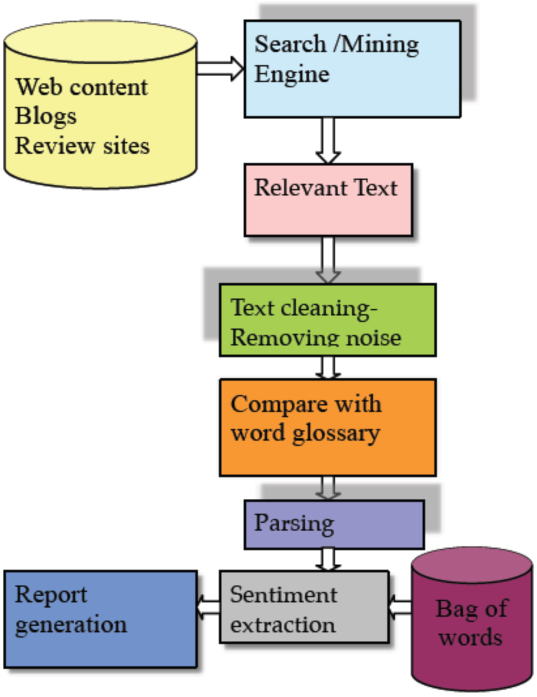
III. RELATED WORK

One of the first attempts in this field was in identifying the genre of texts, for instance subjective genres (Karlgen and Cutting, 1994; Finn et al., 2002). The initial approaches to sentiment detection used linguistic heuristics, explicit list of preselected words and other such techniques that require use of experts' knowledge and may not yield the best possible results in all cases as pointed out in Bo Pang et ., 2002. The first attempt to automate the task of sentiment classification was seen in the work of Turney(2002). He used the mutual information between a document phrase and the words "excellent" and "poor" as a metric for classification. The mutual information was determined on the basis of several techniques are used for the opinion mining tasks. To extract opinions, machine learning and lexical pattern extraction were used by many researchers. In 2002, Turney introduced the results of review classification by considering the algebraic sum of the orientation of terms as respective of the orientation of the documents but more sophisticated approaches are introduced by focusing on some specific tasks: finding the sentiment of words by Hatzivassiloglou , Wibe , Riloff et al , Whitelaw et al , Dave et al. subjective expression by Wilson et al

Pang and Lee were the first to apply machine learning techniques to text classification problem. During feature selection, they used the Bag-of Words approach and extracted nearly 16000 features. For learning they used Naive bayes, Maximum Entropy and Support Vector Machine Algorithm under a 3 Fold cross validation evaluation, good observations using bigrams (2 word combinations), POS tagging etc. Lee again extended their previous work in which they extracted only the subjective sentences by filtering the nonsubjective ones. Here they extended the data set to 2000 equally distributed reviews and made it standard. They have obtained comparable performances over the previous one. Konig and Brill used a hybrid classifier which works in two steps; in the first phase they used a pattern based classifier and if the document is not classifiable at first phase, it is sent to general learning based classifier at second step . In India many companies like Valuepitch Interactive, Pinstorm are doing sentiment analysis and they have many clients across India who want to keep track of the sentiments about their products.

IV. PROPOSED SYSTEM ARCHITECTURE

The proposed system architecture is shown in the following figure:



The steps involved in extraction of sentiment are:

A. Extraction of Data

Extraction of data involves the extraction of historical data and textual data related to the performance of the firm, i.e. news articles etc. A web crawler is used to extract all the required web pages. Retrieving historical stock values is easy, however, extraction of textual data is more involved due to the differences in the webpage design and structure.

B. Text Preprocessing

A text preprocessor is tailor made for a specific case. This involves cleaning HTML and other web documents by retaining only the relevant information.

C. Extracting the Sentiment

The two main methods available for extraction of sentiment : Lexical analysis and machine learning

Lexical Analysis involves :
- Initializing the polarity score p = 0
- Tokenizing the text and for each token :
• Check if it is present in the dictionary
• If present, add the sentiment weight w
to p :- p = p + w (w may be a positive
or negative quantity)
-Check score p,
• If p > threshold, then post is positive
• If p < threshold, then post is negative

V. DATA COLLECTION

The analysis uses two different datasets : A large number of news articles and expert opinions, a historical dataset containing stock values. The stock values are over last 5 years ie. from Jan 2010 to Jan 2016. The historical stock values are from https:// in.finance.yahoo.com

The list of stock values range from Jan 2010 to Jan 2016 of INFOSYS in NIFTY. The news articles are extracted from moneycontrol website as it contains relevant news regarding individual stocks. A web Relevant Text Search /Mining Engine Text cleaning- Removing noise Compare with word glossary Parsing Sentiment extraction Bag of words Report generation scraper was built in python to extract the web pages and calculate the sentiment.

VI. EXPERIMENT

The sentiment was extracted from articles using lexical analysis. Articles themselves were extracted from moneycontrol website. A total of 2100 articles regarding infosys, published between January 2010 and January 2016, were extracted and cleaned. The articles were then analyzed and scored.

Score was calculated according to the following formula:

$$\text{Score} = \frac{\Sigma Positive\ scores - \Sigma Negative\ scores}{\Sigma Positive\ scores + \Sigma Negative\ score}$$

where,

$$\Sigma Positive\ scores = Sum\ of\ all\ scores\ of\ positive\ words$$

$$\Sigma Negative\ scores = Sum\ of\ all\ scores\ of\ negative\ words$$

The computed values are then plotted according to the formula,

$$Predicted\ Stock\ Value_i = Predicted\ Stock\ Value_{i-1} \times (1 + \alpha \cdot Score_i)$$

Where,

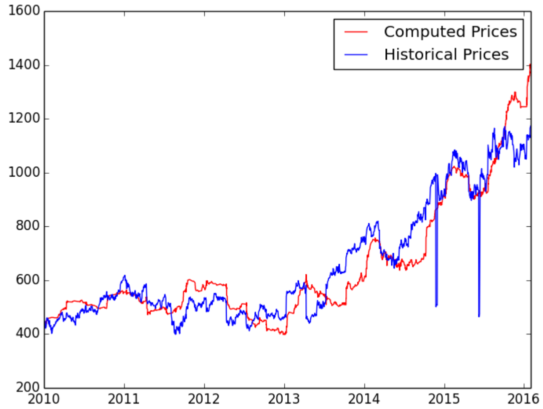
$$Value_i = value\ on\ i^{th}\ day$$

$$\alpha = Scaling\ factor$$

$$Predicted\ Stock\ Value_0 = Historical\ Stock\ Value_0$$

VII. RESULT

The correlation between historical stock prices and prices calculated using sentiment analysis is shown in the graph:



VIII. CONCLUSION AND FUTURE WORK

It can be concluded that market sentiment both drives and is driven by actual stock prices. Complex models can further be developed that incorporates sentiment analysis as an intermediary step. These models using machine learning can be trained to predict how the market behaves given the current sentiment of the market.

REFERENCES

[1]. CaslavBozi, DetlefSeese: Neural Networks for Sentiment Detection in Financial. JEL Codes: C45, D83, an G17, Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology (KIT).

[2]. Liang, X.: Impacts of Internet Stock News on Stock Markets Based on Neural Networks. ISNN 2005, LNCS 3497, pp. 897--03, Springer-Verlag, Berlin Heidelberg, 2005.

[3]. Liang, X., Chen, R..: Mining Stock News in Cyber world Based on Natural Language Processing and Neural Networks. ICNN&B '05, pp.13--15.

[4]. VivekSehgal., Charles Son.: SOPS: Stock Prediction using Web Sentiment. IEEE, DOI 10.1109,2007.

[5]. KhurshidAhmad, Yousif Almas.: Visualizing Sentiments in Financial Texts.

[6]. Yang Gao, Li Zhou, Yong Zhang, Chunxiao Xing.: Sentiment Classification for Stock News. 978-1-4244-9142-1/10/2010 IEEE.

[7]. SusanneGlissman, Ignacio Terrizzano, Ana Lelescu, Jorge Sanz.: Systematic Web Data Mining with Business Architecture to Enhance Business Assessment Services. Annual SRII Global Conference, 9 78-0-7695-4371-0/11, 2011 IEEE DOI 10.1109.

[8]. Hailiang Chen, Prabuddha De, Yu (Jeffrey) Hu, Byoung-HyounHwang.: Sentiment Revealed in Social Media and Effect on the Stock Market. STATISTICAL SIGNAL PROCESSING WORKSHOP,IEEE.

[9]. Yong LI, JianWANG: Factors on IPO under-pricing based on Behavioral Finance Theory: Evidence from China. 978-1-4577-0536-6/11/2011 IEEE.

[10]. Kaihui Zhang1, Lei Li2, Peng Li3,and Wenda: Stock Trend Forecasting Method Based on Sentiment Analysis and System Similarity Model. 11. Binoy.B.Nair, V.P Mohandas, N. R. Sakthivel.: A Decision Tree- Rough Set Hybrid System for Stock Market Trend Prediction. International Journal of Computer Applications (0975 – 8887), Volume 6– No.9, September 2010

[11]. Paul D,Yoo., Maria H., Kim, Tony Jan.: Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation.IEEE,2005.

[12]. http://www.cs.uic.edu/~liub/FBS/sentimentanalysis. html