

Analyzing stock price changes using event related Twitter feeds

Satyabrata Aich¹, Hee-Cheol Kim¹, Mangal Sain², Bijay Bhaskar Deo³

¹Department of computer engineering, Inje University, S. Korea

²Department of Computer Engineering, Dongseo University, S. Korea

³Department of computer science, Syracuse University

satyabrataaich@gmail.com, heeki@inje.ac.kr, mangalsain1@gmail.com, deo.developer@gmail.com

Abstract:- For many years twitter has been the key source for providing the sentiments of the people in the real time, which are helpful for analyzing the stock market. But after certain event happened the sentiments of people became more predominant on the stock price. This paper describes about the stock prices changes over the event related day wise tweet sentiment score. We have over 200000 tweets over the period of 5 days. We have analyzed the stock market values collected from yahoo finance. We have done the keyword analysis of tweets to find the frequent keywords used during the period about the event and also we have seen the trend of keyword used over the period. After all the analysis and observation we found some positive correlation between the stock prices changes and the average sentiment score.

Keywords:- Sentiment analyses, stock market, twitter, sentiment score, tweets

I. INTRODUCTION

The prediction of stock market has been an active research area for many years. The discovery of new technologies makes it more interesting. Although lot of study has been done on the prediction of stock market, so far no clear method has been established to predict the stock. These days all the researchers and the leading financial institutions are looking for maximizing the returns with minimum risk. So to minimize the risk it is necessary to know the factors that affect the stock market. The most important factors that affect stock market are classified into three categories: fundamental factors, technical factors and the market sentiment. For long time it has been contemplated that market sentiment is an ambiguous factor and its effects are huge and biased for the prediction of stock [1]. The efficient market hypothesis (EMH) advocates that social media sentiments helps to understand the stock market better and are able to help in maximizing the return. So sentiment analysis or the opinion mining, which extracts the information's from the sentiments will helps to understand the stock market better.

According to Bing Liu [2] Sentiment analysis, also called opinion mining, is the field of study that analyze people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. These days most of the big companies are using opinion mining as the key strategic tool for predicting the market [3, 4, and 5].

Twitter, the most famous microblogging site on which people posts their opinion on the real time has become popular for predicting the stock market. Lots of work has been done by analyzing the sentiment of the people for the prediction of stock market. However they have used different technique for analyzing the stock market. Less work has been done on the event related twitter feeds affecting the stock prices. Till now no work has been done to find the relationship between sentiment score and event related twitter feeds. So in this paper we have analyzed the sentiments of the company related tweets after an event happened and based on the polarity of the sentiments with a score ranging from +1 to -1, we calculated the sentiment score and then tried to correlate the sentiment score with the changes in the stock price for a certain period. We have also analyzed the most frequent keywords used in the tweets for that period and showed the word matrix and also the trend of the frequent keywords over that period.

The structure of the paper is as follows: Section II presents the past work related to the twitter sentiment and stock market. Section III describes the details about the data collection and proposed technique. Section IV describes about the data analysis and discussion. Section V discusses about the conclusion and future work.

II. RELATED WORK

Chen & Lazer [6] has investigated the relationship between the movement of stock market and twitter feeds. They identified several features affecting the sentiment values and used classification and regression strategy to predict the movement of stock market.

Zhang [7] has found out the correlation between twitter sentiment and stock prices and the words in the tweets correlate to price changes in stock prices. They compared the result of three machine learning techniques and the effectiveness of the techniques on the correlation. They correlate the twitter sentiment score and stock price by gathering the stock data minute by minute for a day.

Bollen et al [8] has tried to find out the behaviour of the stock market by considering the mood of the people on the twitter. They found out the moods by analyzing the twitter text for a certain period and observed the changes in mood state for that period and by using that data they studied the prediction of stock market.

Mittal & Goel [9] has also studied the stock market prediction based on the mood state generated from twitter feeds. Their approach is also very similar to the bollen et al's approach mentioned above.

Rao & Srivastava [10] have identified the relationship between the tweet board literature (like bullishness, volume, agreement etc) with the financial market instruments (like volatility, trading volume and stock prices). They found that the public mood has a greater influence on the movement of stock indices.

Smailović et al [11] attempted to find out whether twitter feeds about the companies and their products are the reliable source for the movement of stock prices. They used one new indicator, named positive sentiment probability to be used in the predictive sentiment analysis of finance. They also used the classification techniques as an improvement in the predictive power for stock market application.

Ranco et al [12] presented some significant evidence for measuring the dependence between the sentiment in tweets about the companies and the stock price returns. They observed the market behaviours after identifying the events that makes the increase amount of posts related to the event among the twitter users.

Mao et al [13] investigated whether the number of tweets for a particular day has any correlation with the stock price and traded volume for that day. They examined it at three different levels such as stock market, industry sector and individual company stocks. They found that the correlation is positive between them.

All of the above mentioned work motivated us to propose our work which tried to find out the correlation between the sentiment score and stock price changes for a certain period after an event.

III. DATA COLLECTION AND PROPOSED TECHNIQUE

A. Data Collection

We have collected the tweets using the twitter streaming API. We have used python 3.5.2 programming language for the scripts. We used "Tweepy" an easy to use python library for accessing the twitter API. The tweets are filtered based on the keywords. In our case we have collected the tweets about the company "Samsung" and its product. The tweets are saved in sqlite database. We have collected more than 200000 tweets. As we mentioned before we collected the tweets after the event that is related to the galaxy note 7. We have also collected the stock related data about the Samsung Company for the above mentioned period.

B. Proposed Technique

Python has a library named "TextBlob" for processing the textual data. It has all the features required for the sentiment analysis of the tweets. It has an integrated Wordnet Lexicon. We developed our script for finding sentiment score of a tweet ranging from +1 to -1. +1 refers to highly positive tweets and -1 refers to the highly negative tweets. After getting the

individual score we calculated the average score of the tweets for a particular day by using SQL scripts. We have also carried out the keyword analysis for better understanding about the event and for improved correlation. The analysis has been done using "VOYANT [14]. VOYANT Tools is a web-based reading and analysis environment for digital texts. The flow chart of the proposed technique is mentioned below in Fig. 1.

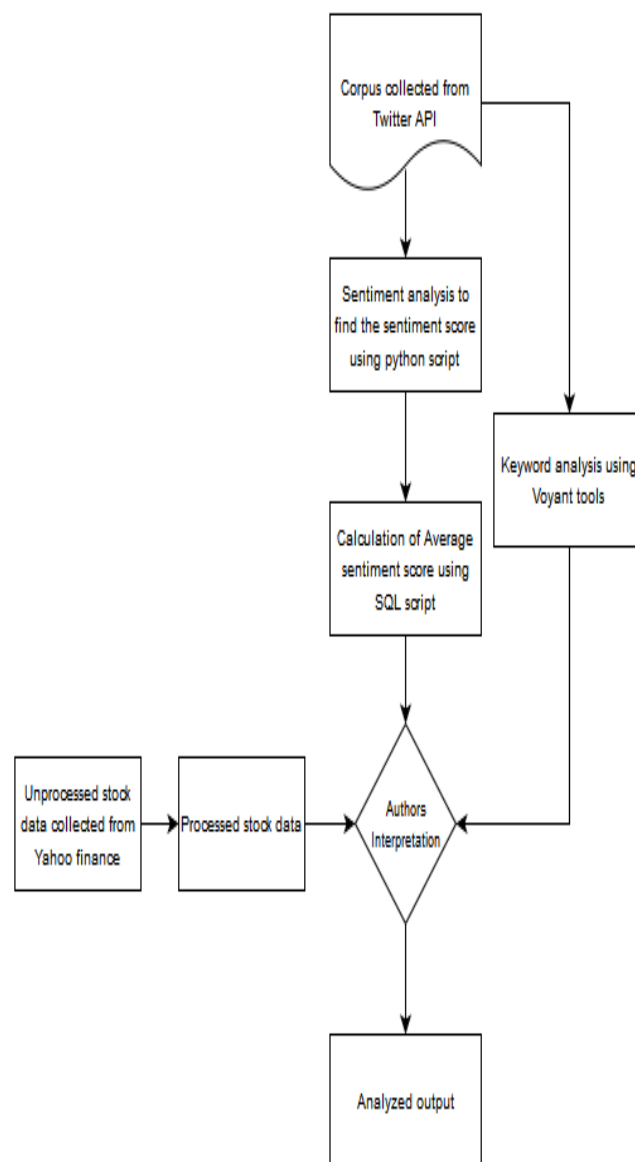


Figure 1. Flow chart of proposed technique

IV. DATA ANALYSIS AND DISCUSSION

A. Sentimental Analysis

As we mentioned before according to our python script we have able to calculate the sentiment score of the each tweets on a day. The score ranges from +1 to -1. Since we have tweets that have positive value as well as negative value and

TABLE1. Day wise average sentiment score

Day	Sentiment Score
17-10-2016	0.1005
18-10-2016	0.1157
19-10-2016	0.0982
20-10-2016	0.1098
21-10-2016	0.0951

We have done this analysis to see which word is frequently used in the tweets in the above mentioned time period. The summary of the analysis said that this corpus has 1 document with 4,005,238 total words and 235,719 unique word forms. Most frequent words in the Corpus: samsung (157634); galaxy (113917); note (59692); iphone (15732); note7 (14713).The keyword frequency diagram is shown in Fig. 2.

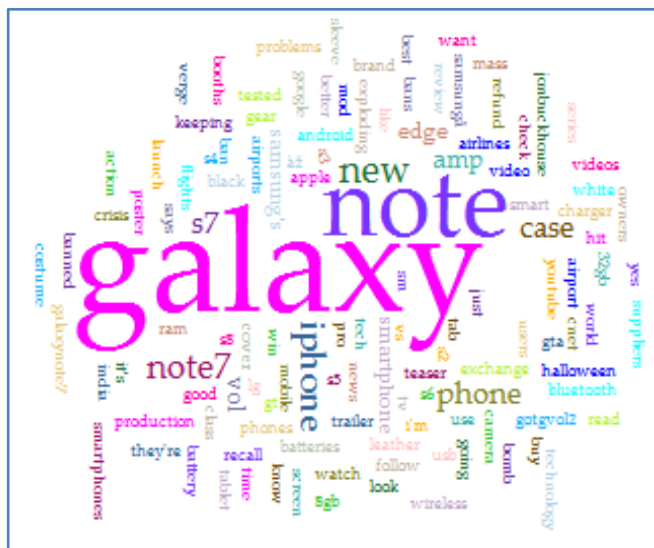


Figure 2. Frequency of keywords mentioned in the tweets

It is clear from the analysis that since the galaxy note 7 issue was a big event in Samsung recently, these words such as galaxy, note7, iphone etc. are frequently mentioned in the tweets. Fig. 3 shows the trend of the keywords used over the time period. From the diagram it is shown that the word galaxy has an upward trend and slowly it goes downward over the time period, where as the trend of iphone doesn't vary much over the time period.

From the Fig. 2 and Fig. 3, we can conclude that after certain event happened twitter used to receive lots of tweets about the event. In our case we took the Samsung galaxy 7 issue as an event, because of that the words like Samsung and galaxy has a high trend in the documents.

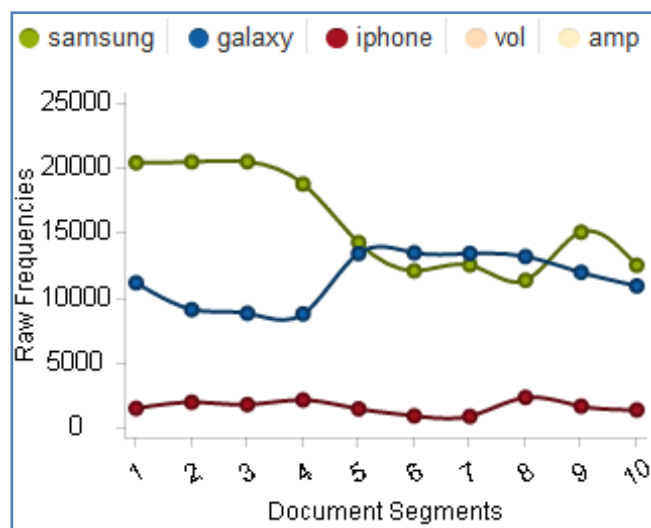


Figure 3. Trend of keywords mentioned in the tweets

We have collected the one month data from yahoo finance [14]. Fig. 4 shows the stock price movement over one month. The opening and closing price of the stock are represented by blue and red line respectively. We can see from the Fig. 4 that after the galaxy note 7 issues came into picture on the 2nd week of October, after that Samsung immediately stopped the production of it, the stock prices has a big fall on that period and its impact was so strong that the stock price could not able to recover fully till now although the trend of the curve is going upward. So it is clear from the Fig. 4 that an event or issue related to company has a major impact on stock price. To make it more clear we have related the sentiments with respected to stock price movement that is shown in the Fig. 5.

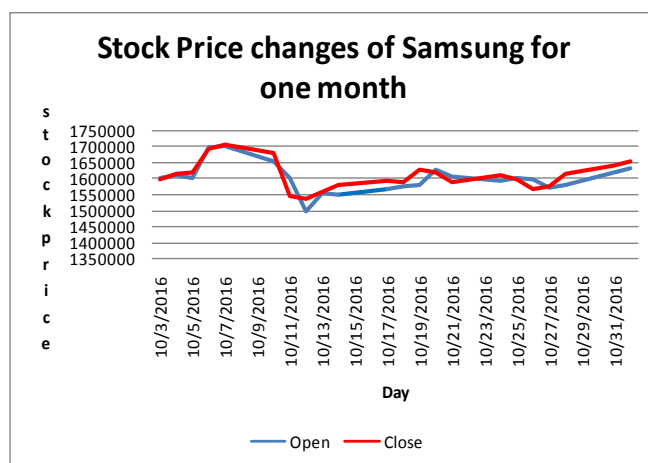


Figure 4. Stock price changes of Samsung for one month

Fig. 5 below shows the stock price changes with respect to sentiment score for Samsung over the period of five days, starting from 17th to 21st October. The closing and opening price of the stock for each day are represented by the violet and green line respectively. The sentiment score represented by the red line.

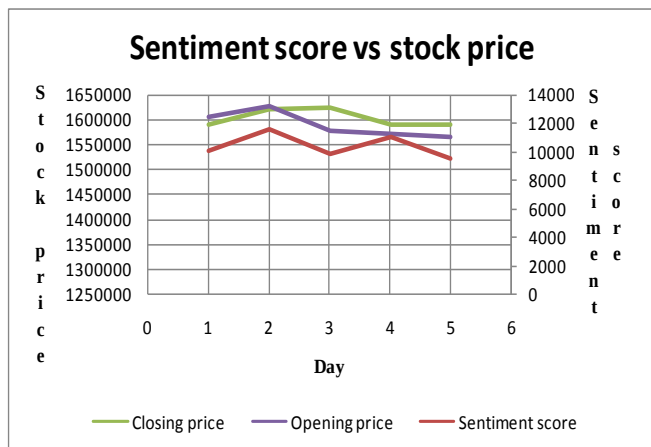


Figure 5. Stock price change with sentiment score

V. CONCLUSION AND FUTURE WORK

In this paper we have observed the stock prices changes with respect to the sentiment score of the event related tweets. We found some positive correlation between the data sets. We have seen from the keyword analysis that the keywords mentioned in the tweets are more related to the event. Even the trend of keywords shows that how the sentiment has a big impact these days on the market. Although our analysis period includes tweet from 5 days but since the event, which is issue about galaxy note 7 is so big so its impact has been active for a long time on Samsung. In the future we will collect more tweets for a long period of time and analyze more deeply into with other parameters. We will also use different classifiers to classify the tweets and also we will use other technique such as deep learning, topic model to see the correlation between twitter feeds and stock prices movement.

REFERENCES

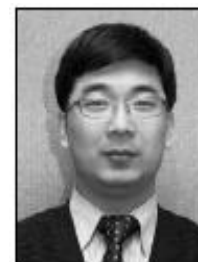
- [1] D. Harper, Forces that Move Stock Prices. (2010) [Online] Available at: <http://www.investopedia.com/articles/basics/04/100804.asp> [Accessed on 27th October 2016]
- [2] B. Liu, Sentiment analysis and opinion mining, Morgan and Claypool publishers, 2012
- [3] S.Das and M.Chen, "Yahoo! for Amazon: Extracting market sentiment from stock message boards", Proceedings of APFA-2001.
- [4] S.Das and M.Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web", *Management Science*, 53(9): 1375-1388, 2007
- [5] Y.Chen and J.Xie, "Online Consumer Review: Word of Mouth as a New Element of Marketing Communication Mix", *Management Science*, March 2008, vol 54, no 3, 477-491, 2008.
- [6] R. Chen and M. Lazer, "Sentiment analysis of twitter feeds for the prediction of stock market movement", Stanford. edu. Retrieved January. 2013, 25Th Jan 2013
- [7] L. Zhang, "Sentiment Analysis on Twitter with Stock Price and Significant Keyword Correlation", The University of Texas at Austin, 2013
- [8] J. Bollen, H. Mao and X. Zeng, "Twitter mood predicts the stock market", *Journal of Computational Science*, 2(1), pp.1-8, 2011
- [9] A. Mittal and A. Goel, "Stock prediction using twitter sentiment analysis", Stanford University, CS229 (2011 <http://cs229.stanford.edu/proj2011/GoelMittalStockMarketPredictionUsingTwitterSentimentAnalysis.pdf>), 2012
- [10] T. Rao and S. Srivastava, "Analyzing stock market movements using twitter sentiment analysis", *In Proceedings of the 2012 International*

Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 119-123, IEEE Computer Society, 2012

- [11] J. Smailović, M. Grčar, N. Lavrač and M. Žnidaršič, "Predictive sentiment analysis of tweets: A stock market application, In Human-Computer Interaction and Knowledge Discovery in Complex, Unstructured, Big Data", pp. 77-88, Springer Berlin Heidelberg, 2013
- [12] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The effects of Twitter sentiment on stock price returns", *PloS one*, 10(9), p.e0138441, 2015
- [13] Y. Mao, W. Wei, B. Wang and B. Liu, "Correlating S&P 500 stocks with Twitter data", In Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research, pp. 69-72, ACM.(2012)
- [14] <https://voyant-tools.org>
- [15] <https://finance.yahoo.com/quote/005930.KS/history?p=005930.KS> [Accessed on 1st November 2016]



Satyabrata Aich is working as a researcher in the field of computer engineering. He has over four years of teaching, research and industry experience in India and abroad. He has published many research papers in journals and conferences in the realms of Supply Chain Management and data analytics. His research interests are natural language processing, Machine learning, supply chain management, data mining.



Hee-Cheol Kim received his BSc at Department of Mathematics, MSc at Department of Computer Science in SoGang University in Korea, and PhD at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. His primary concerns are u-Healthcare, smart home technology, e-learning and Human Computer Interaction (HCI). He is associate professor at Department of Computer Engineering, Inje University in Korea. He has published many research papers in journals and conferences in the realms of HCI and CSCW.



and a TPC member in several international conferences

Mangal Sain received the M.Sc. degree in computer application from India in 2003 and the Ph.D. degree in computer science in 2011. Since 2012, he has been an Assistant Professor with the Department of Computer Engineering, Dongseo University, South Korea. His research interest includes wireless sensor network, cloud computing, Internet of Things, embedded systems, and middleware. He has authored over 30 international publications including journals and international conferences. He is a member of TIIS



Bijaybhaskar deo Graduated from Syracuse University with Masters in Computer Science. He has 10 years of professional experience in IT Industry involving various phases of Software development including design, development, production support of robust, scalable products and applications on a variety of software platforms. Current, A product engineer on web analytics involving Big data areas.