Opinion Mining and Sentiment Analysis

Rushlene Kaur Bakshi
Department of Computer Science and Engineering, Guru Tegh Bahadur Institute of Technology, G-8 Area, Rajouri Garden, New Delhi-110064, INDIA
Email ID: rushlenebakshi@gmail.com

Navneet Kaur
Department of Computer Science and Engineering, Guru Tegh Bahadur Institute of Technology, G-8 Area, Rajouri Garden, New Delhi-110064, INDIA
Email ID: nkaur551@gmail.com

Ravneet Kaur
Department of Computer Science and Engineering, Guru Tegh Bahadur Institute of Technology, G-8 Area, Rajouri Garden, New Delhi-110064, INDIA
Email ID: ravneetkaur195@gmail.com

Gurpreet Kaur
Department of Computer Science and Engineering, Guru Tegh Bahadur Institute of Technology, G-8 Area, Rajouri Garden, New Delhi-110064, INDIA
Email ID: gurpreet_gk@live.com

***Abstract*-In the recent years, micro blogging platforms like Twitter have become instrumental in gauging public mood. So it makes it a feasible predictive strategy to speculate the rise and fall of stock prices. This paper aims to undertake a stepwise methodology to determine the effects of an average person's tweets over fluctuation of stock prices of a multinational firm called Samsung Electronics Ltd. It involves extracting tweets from twitter, data cleansing and application of a suitable algorithm in order to get the adequate sentiment analysis. The vast impact created by twitter data feed has been greatly studied in this paper. Attempts have been made to design an algorithm which works well analysing the positive, negative and neutral tweets.**

Keywords: Opinion Mining; Sentiment Analysis; Twitter; Natural Language Processing; Micro Blogging

I. INTRODUCTION

A. Opinion Mining

Human decision making is extensively influenced by the assessment or judgement of others. Before making any move, customers tend to gather as much information as possible about the product they want to buy. The investors try to analyse and predict the stock market movement of a particular company based on its popularity among its customers before investing their money in its shares. With the advent and development of social media, gathering data for evaluation has become easier and less time consuming. Different platforms like Twitter, Facebook, LinkedIn serve as repositories of useful data in terms of reviews, likes, comments etc.[6] The industrialists, businessmen and manufacturers use these results to scrutinize and maintain their quality standards. Sentiment Analysis, another term for opinion mining, is a language independent technology and is also applied in the study of sociology, law, psychology, politics, management etc. Collectively, analysis of all these reviews and data is an emerging field of research called opinion mining. [1]

B. Natural Language Processing

Natural Language Processing (NLP) is technique is largely being used to decipher human language. It is used to measure how accurately a system can meet its goals qualitatively. It involves tokenization and machine level processing .This may either involve classification of dataset using supervised learning or finding hidden` patterns in unstructured data using unsupervised learning [11, 14]. A large sentiment based lexicon forms a part of the NLP domain to classify human emotions.

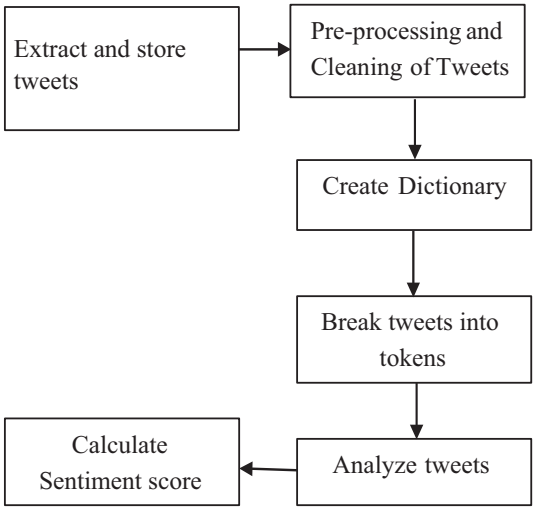
II. RELATED WORK

Study of opinion mining has been applied to a plethora of areas. The most common use being analysis of products or services of a company in order to improve it according to the user's need. Not being restricted to this, it is used by political parties to get an insight of their impression on the public and have a reality check for the forthcoming elections. [1, 4] Government policy making bodies use sentiment analysis tools [12] to judge the public mood with respect to a new policy or reformation of an old one. If the public doesn't support the idea, the government works to mould it accordingly. In the arena of trading and marketing, opinion mining is immensely important as its effects can make or break the entire business. Market research techniques like Voice of Customer (VOC) [13] use sentiment analysis in order to analyse the customer's expectations or grievances. [10] E-commerce websites and manufactures provide their customers online feedback submission option which they later review using opinion mining. Based on the results of these, stock market predictability for a company or product can be done. It will help investors to take wise decisions about their investment in that company's shares. [1] Work in this paper focusses on unsupervised learning using sentiment-lexicon based approach. Although Machine learning technique is a better method, it requires a trained dataset, which is extremely time consuming. Instead, an unstructured data set is used here and appropriate

978-9-3805-4421-2/16/$31.00 ©2016 IEEE

sentiment score is detected after gathering the polarity of words. [9]

General approaches used for opinion mining are - Sentence level, Document level and Feature level. Sentence level approach is being used here. [8,9]



:

Fig.1.Flowchart of system

Firstly, take the tweets from twitter and store them into MS Access database.

TABLE I. DATABASE TABLE

Table Name	Field Name	Data Type
Analyse Score	Tweet	Text
	Score	Number
	User Name	Text
Dictionary	Word	Text
	Priority	Number

Make a dictionary for positive, negative and neutral words and assign priorities as:

Positive Words=1
Negative Words=-1
Neutral Words=0
For eg. Consider the following table:
Make a dictionary for positive, negative and neutral words and assign priorities as:

Positive Words=1
Negative Words=-1
Neutral Words=0
For eg. Consider the following table:

TABLE II. DICTIONARY TABLE

Word	Priority
Abort	-1
Accuse	-1
Stable	1
Success	1

Now, pre-processing will be done by retrieving the tweets from database and perform cleansing operations by removing unwanted characters from each tweet. Split the complete tweet into individual words and store them into an array. Compare each word in dictionary and if word is found, assign the corresponding priority to it. Calculate the Sentiment Score [3].

TABLE III. SENTIMENT SCORE DATABASE TABLE

User	Tweet	Sentiment Score
Samsung	Screen failure in less than 2 weeks-s6	-1
Samsung	finally the phone from samsung I've been waiting for	0
Samsung	Better camera, Faster processor.	2

If
Sentiment Score >= 1, then, tweet is positive;
Sentiment Score <= -1, then, tweet is negative;
Sentiment Score = 0, then, tweet is neutral.
For example:

(1) The best phone on the market. It has an excellent screen, fast processor and wireless charging.
(2) Samsung refused to repair mobile phone under warranty
(3) Bought Samsung a5, I was facing audio problem on my Samsung device after I purchased it.

These sentences portray varied opinions about Samsung Electronics. Sentence (1) expresses a positive opinion about Samsung phone, sentence (2) and (3) expresses negative opinion about it. This can be done by calculating the sentiment score of the complete sentences like in sentence (1) best, excellent and fast are positive words so sentiment score will be 3/3=1 that means given sentence is positive. Similarly, sentence (2) and (3) are negative sentences.

III. ALGORITHM FOR SENTIMENT ANALYSIS

Problem: Given a list of tweets, return sentiment score for each tweet.

Inputs: Enter a tweet from twitter for analysis

Outputs: Sentiment score for each tweet

Algorithm:

sentimentAnalysis (string tweet)
1. Initialize count=0, size and i=0

2. Break the tweet into words
String Tokenizer st = new StringTokenizer (tweet, “”)
3. Count the no. of words in tweet size = st.CountTokens() and declare two array words and new_words of size ‘size’
4. Store each word of tweet into array words.
Repeat step 4(a) until st.hasMoreTokens()
4(a). words[i] = st.nextToken()
i++
Initialize i=0
5. Create database connection
Class.forName (“sun.jdbc.odbc.JdbcOdbcDriver”)
Con=DriverManager.getConnection(“jdbc:odbc: tweet”, “”, “”)
6. Repeat step 7 to 9 until words.length
7. new_words [i] = words[i]
replaceAll(“[‘.!;:@/?><#$%^*&~’]”, ””)
(Remove unwanted characters from tweet)
8. Create query stat = Con.createStatement() res = stat.executeQuery(“select * from word where word = ‘”+ new_words[i]+” ’ ”)
9. Repeat step 10 until res.next()
10. Count=count+res.getInt(“Priority”)
11. End loop
12. Calculate Sentiment Score using

Sentiment Score =
N(Positive Terms) - N(Negative Terms
N(PositiveTerms) + N(Negative Terms)

13. Calculate Error & Accuracy Percentage using
Error = Actual Score – Calculate Score
Error % = (Error/actual score)*100
Accuracy % = 100 – Error%
Here, the actual score is human predicted score and calculated score is software calculated.
14. End

Here, extraction of tweets from twitter is done using Twitter4j library and string tokenizer is used to break each tweet into individual words. Store all these words in an array. Create JDBC-ODBC connection for storing these tweets and remove all unwanted characters and inconsistent words from each tweet and store them in a new array. Match the words in the new array with the given dictionary and set priority according to sentiment score. Plot a graph of sentiment score v/s tweets with sentiment score on y axis and tweets on x axis. Calculate the error and accuracy percentage for the data used.

The time complexity of this algorithm is O(m*n).

IV. RESULT

The purpose of this research is to design an algorithm that can efficiently compute the sentiments of data coming from twitter. The algorithm efficiency is measured in terms of accuracy rate and time complexity. Tweets are analyzed with an accuracy of 80.6% and its time complexity is O(m*n).

TABLE IV. RESULT TABLE

Positive Tweets	Negative Tweets	Sentiment Score	Score%
27	29	0.0357	0.06
29	33	6.45	0.104
26	38	18.75	0.29

Table 4 shows the count of positive and negative tweets and their sentiment score using the formula:

Sentiment score =

N(Positive Terms) - N(Negative Terms

N(PositiveTerms) + N(Negative Terms)

V. CONCLUSION

Sentiment Analysis is widely being used in a lot of applications today. With ongoing research work in the field, its use is increasing with each passing day. We have carried out an overview of the current state-of-art facilities used to predict nature of tweets and also to gather share prices in the market.
This paper depicts the use of a new algorithm to give accurate results The method used for analysis performs well in both speed and time hence can be applied to larger datasets in the future. The result gathered is in correspondence with slight deviation from current share price. This research mainly concentrates to perform sentiment analysis quickly and in an efficient manner. Though in this approach the analysis was restricted to a single company, it can be applied to stocks of other companies also.

VI. CHALLENGES

- To detect spam messages;
- Application of opinion mining on abbreviations;
- Creating a rich lexicon database;
- Identification of bi polar, sarcastic sentences and interrogative sentences. [7]
- Judgement of context [7]
- Appropriately resolving co-reference in the usage of nouns and pronouns
- Identification of fake reviews [1,2,5]

VII. FUTURE SCOPE

The credibility of the tweets being analysed can be checked by taking reviews from other websites and sources and then comparing the two. The sarcastic element would be further reduced in the future by better processing.

REFERENCES

Journal References

[1] Gautami Tripathi1 and Naganna S2, "Opinion Mining: A Review", International Journal of Information & Computation Technology

[2] Haseena Rahmath P, "Opinion Mining and Sentiment Analysis - Challenges and Applications", International Journal of Application or Innovation in Engineering & Management (IJAIEM), Volume 3, Issue 5, May 2014

[3] DongSung Kim2 and Jong Woo Kim, "Public Opinion Mining on Social Media: A Case Study of Twitter Opinion on Nuclear Power1", Advanced Science and Technology Letters,Vol.51 (CESCUBE 2014), pp.224-228.

[4] Diana Maynard and Adam Funk, "Automatic detection of political opinions in Tweets",ESWC'11 Proceedings of the 8th international conference on The Semantic Web ,Pages 88-99

[5] Nidhi R. Sharma, Prof. Vidya D. Chitre, "Opinion Mining, Analysis and its Challenges", International Journal of Innovations & Advancement in Computer Science, Volume 3, Issue 1, April 2014

[6] Johan Bollen, Huina Mao, Xiao-Jun Zeng, "Twitter mood predicts the stock market", Journal of Computational Science, Vol. 2, 2011, pp1-8

[7] S. ChandraKala, C. Sindhu, "Opinion Mining And Sentiment Classification : A Survey", ICTACT Journal on Soft Computing, Volume 3, Issue 1, October 2012

[8] Walaa Medhat, Ahmed Hassan, Hoda Korashy, "Sentiment Analysis Algorithms and applications : A survey", Ain Shams Engineering Journal, 2014

[9] G. Vinodhini, RM. Chandrasekaran, "Sentiment Analysis and Opinion Mining : A Survey", International Journal of Advanced Research in Computer Science and Software Engineering,Volume 2, Issue 6,, June 2012

[10] Rushabh Shah and Bhoomit Patel, "Procedure of Opinion Mining and Sentiment Analysis : A Study", International Journal of Current Engineering and Technology, Vol. 4, No. 6, December 2014

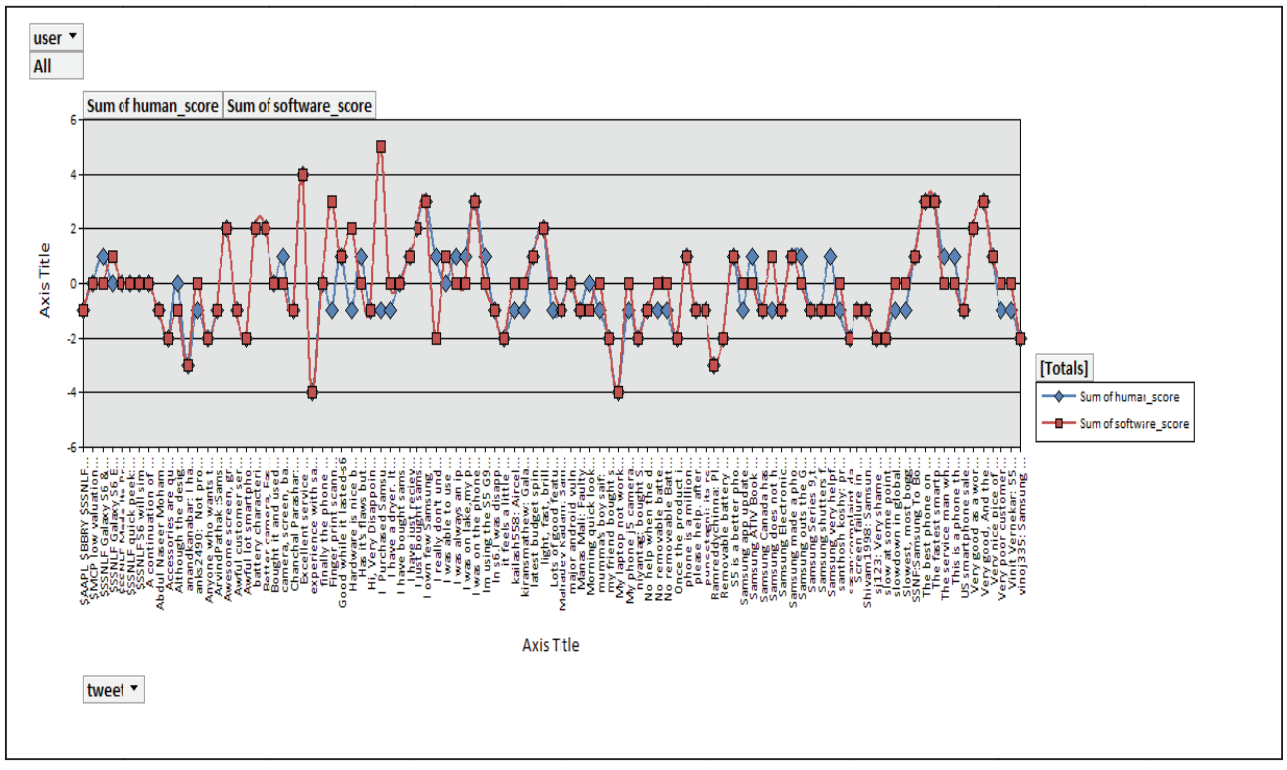
[11] Bing Liu, "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012

[12] G.Angulakshmi & Dr.R.ManickaChezian "An Analysis on Opinion Mining: Techniques and Tools", International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 7, July 2014.

[13] Mining Hu and Bing Liu, "Mining and summarizing customer reviews". Proceedings of the 10th ACM SIGKDD International conference on knowledge discovery and data mining, 2004.

[14] T. Zagibalov and J. Carroll, "Unsupervised classification of sentiment and objectivity in Chinese text", In Proceedings of IJCNLP 2008, Hyderabad, India, January 2008

X. APPENDIX



Graph 1: It shows the sentiment score across each tweet with number of tweets on x-axis and users posting tweets on y-axis.