Enhancement of Sentiment Analysis on Twitter

A.Anantha Lakshmi[#1],M.Iswarya[#1],J.N.Maria Boncy[#1],J.Parameshwari[#1]& S.Hariharan[#2]
Student[#1],Associate Professor[#2]
Department of computer Science and Engineering
TRP Engineering College, Irungalur, Trichy, Tamil Nadu, India
ananthiayyanar95@gmail.com[1], iswarya.abinayam@gmail.com[1], jn.mariaboncy@gmail.com[1],
jothiparameshwari@gmail.com[1],
mailtos.hariharan@gmail.com[2]

***Abstract*: In this paper, we analyze social media data. Social media analytics is the practice of gathering data from blogs and social media websites and analyzing that data to make business decisions. The most common use of social media analytics is to mine customer sentiment in order to support marketing and customer service activities. And then we take twitter big data to predict named entity. In this work, it is aimed to reduce the Twitter user's effort to access to the tweet carrying the information of interest. To this aim, a tweet recommendation method under a user interest model generated via named entities is presented. To achieve our goal, Hybrid Segmentation is generated via named entities extracted from user's followers' and user's own posts. And extend our approach to analyze short text in tweets.**

.

***Keywords*: Big data, Hybrid Segmentation, Stemming, POS tagging, Map Reduce.**

I. INTRODUCTION

Sentiment analysis has been an active area of research in recent years [1]. This paper introduces sentiment analysis on twitter. It is the classification of the polarity of a given text in the document, sentence or phrase. The goal is determined whether the expressed opinion in the text is positive, negative and neutral. Since microblogs plays a vital role and provides a platform for the interaction for the public issues.In social media the twitter may have a different opinion on different topics. The users tweet their view publicly in twitter. This idea makes our paper more convenient to analyse the sentiment of the people on public issues and gives a clear view of how majority of people react to a particular issue.

To study the sentiment analysis in micro blogs which is in view to analyse the tweets in the form of text from the users. The user's tweets are retrieved as a raw tweets.Then these tweets are analysed in order to classify.This classification makes a way to segregate them according to their particular sentiment they move i.e. positive or negative or neutral. Ultimately providing a clear view of sentiment analysis.

The raw tweets are mainly injected from the twitter which acts as a user interface. The collected tweets are then analyzed by classifying. The classification process in mainly based on how and at what perspective this particular word sticks on to the sentiment of positive or negative or neutral. Then these words are shifted to next step of analysis part.These analysed tweets are finally represented as a chart for a particular product.

The paper is organized as follows. Section presents the introduction on sentiment analysis. The basics on big data it is outlined in section 2 while the related work on sentiment analysis is presented in section 3. Section 4 elaborates the architecture with the modules in detail in section 5. Finally conclusion and future work is presented in section 6.

II. BIG DATA

Big data is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process them using traditional data processing applications. The challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, prevent diseases, combat crime and so on."

Velocity - The term 'velocity' in the context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and the challenges which lie ahead in the path of growth and development.

Variability - This is a factor which can be a problem for those who analyze the data. This refers to the inconsistency which can be shown by the data at times, thus hampering the process of being able to handle and manage the data effectively.

Veracity - The quality of the data being captured can vary greatly. Accuracy of analysis depends on the veracity of the source data.

Complexity - Data management can become a very complex process, especially when large volumes of data come from multiple sources. These data need to be linked, connected and correlated in order to be able to grasp the information that is supposed to be conveyed by these data. This situation, is therefore, termed as the 'complexity' of Big Data.

III. RELATED WORK

Big data can be analyzed with the software tools commonly used as part of advanced analytics disciplines such as predictive analytics that includes Hadoop and related tools such as YARN, MapReduce, Spark, Hive and Pigas. Those technologies form the core of an open source software framework that supports the processing of large and diverse data sets across clustered systems.Big data is difficult to work with using most relational database management systems and desktop statistics and visualization packages, requiring instead "massively parallel software running on tens, hundreds, or even thousands of servers". What is considered "big data" varies depending on the capabilities of the organization managing the set, and on the capabilities of the applications that are traditionally used to process and analyze the data set in its domain [4].

This paper explains about a targeting sentiment in twitter; what is considered to be "Big" today will not be so years ahead. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration.Let us consider an example “read the book”[5] it could be positive in book review but negative as in case of the movie review. In social media in twitter may have a different opinion on different topics from reference [11].

IV. PROPOSED WORK

In sentiment analysis on twitter, the raw tweets are injected from the user interface. Then raw tweets are analyzed by the classifications: filtering, tokenization, stop words.

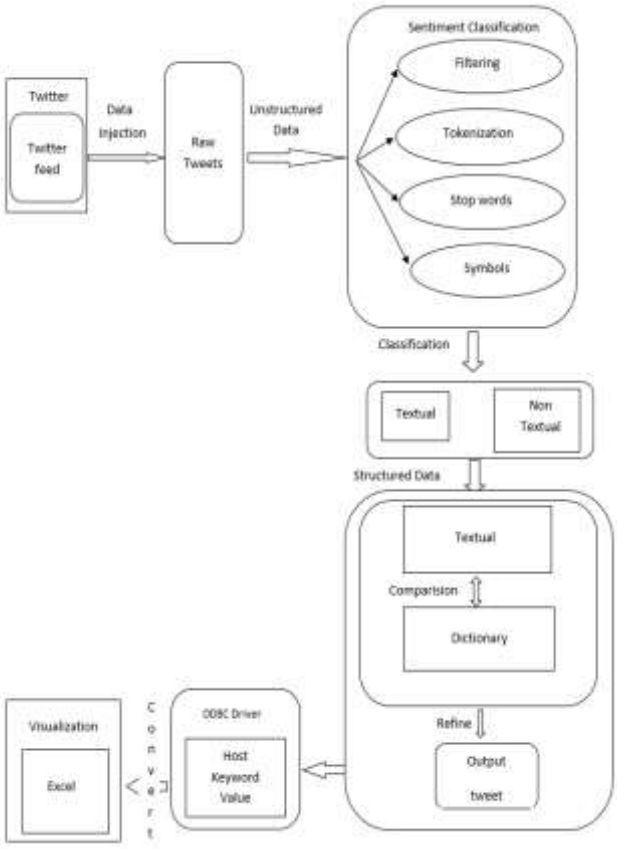

Fig 1.Architecture

In this architecture, a brief overview of sentiment analysis is given in Figure 1.

In order to analyze its sentiment the following steps has been undergone in this paper:
1. Set up a virtual environment.
2. Create a user Interface.
3. Start the server.
4. Collect the Raw tweets for classification.
5. Sentiment classification.

In filtering the special characters (ie) @,#,! Are removed. Stop words analysis will remove the conjunction (is,was). Then the raw tweets are separated into smiley and non-smiley tweets.

The non-smiley tweets are connected to twitter emulator. In emulator, tweets are compared to a dictionary words. Then the results of experiments shows that whether tweets are positive, negative and neutral. The output is represented in a graphical manner.

V. MODULE DESCRIPTION

The module 1 describes about the removal of special characters. Since tweets in the twitter can be twitted using special characters like @,#,! Are removed. The following will be shown in the table 1.

The module 2 describes about the removal of misspelling i.e. The colloquial language of twitting will be removed. The following will be shown in the table 2.

The module 3 describes about the POS tagging , where in POS tagging the sentence will be divided according to their parts of speech as subject, verb, object, adjective, adverb as shown in table 3.

The module 4 describes about the Stemming words preposition will be removed as is, was, these, those, that, as followed by stop words will also be removed as shown in table 4.

TABLE 1
SAMPLE SPECIAL CHARACTERS

special character	@,:,)	@mike fish fair enough.But I have the kindle2 and I think it’s perfect :)	mike fish fair enough.But I have the kindle2 and I think it’s perfect
special character	#,!	Ok,first assessment of #kindle2, it fucking rocks!!!	Ok,first assessment of kindle2 ,it fucking rocks

TABLE 2
MISSPELLING WORDS (SAMPLE)

Misspelling	Thr,without space	Johncmayer is bobby flay joining you	John mayor is bobby flay joining you
Misspelling	Half word	Discreet maths is an inteesting	Discrete maths is an

		subject.	interesting subject.

TABLE 3
POS TAGGING

Pos tagging	Cc	He or /cc she likes skilling	Men like skilling
Pos tagging	Cc	Jean or /cc marylkes singing	Girls likes singing

TABLE 4
STEMMING WORD

Stemming	Are	The boy's car is different colors.	The boy car be differ color.
Stemming	Ty	The catty drinks milk	The cat drinks milk.

VI. CONCLUSION

We designed novel features for use in the classification of tweets in order to develop a system through which informational data may be filtered from the conversations, which are not of much value in the context of searching for immediate information for relief efforts or bystanders to utilize in order to minimize damages.The results of our experiments show that classifying tweets as "positive", "negative" and "neutral" can use solely the proposed features if computing resources are concerned, since the computing power required to process data into featured is immensely decreased in comparison to a BOW feature set which contains a substantially larger number of features.

ACKNOWLEDGEMENT

However, if computing power and time necessary to process incoming Twitter data are not a concern, a combined feature set of the proposed features and BOW-presence approach will maximize overall accuracy.In future work, we can extend our approach implement various classification algorithm to predict the attackers and also eliminate the attackers from twitter datasets. And try this approach to implement in various languages in twitter.

REFERENCES

- Shengnua Liu, Xueqi Cheng, Fuxin Li and Fangtao Li "TASC: topic-adaptive sentiment classification on dynamic tweets " IEEE transactions on Knowledge and Data Engineering,2013.
- Elisa Sarlan, Chayanit Nadam, Shuib Basri. "Twitter sentiment analysis" International Conference On Information Technology and Multimedia, 2014.
- ManjuVenugopalan, Deepa Gupta " Exploring Sentiment Analysis" IEEE transactions on Knowledge and Data Engineering.
- Tiara, Mira KaniaSabariah, Veronikha Effendi, School of Computing, Telkom University Bandung, Indonesia 2015 3rd International Conference on Information and Communication Technology (ICoICT) "Sentiment Analysis on Twitter Using the Combination of Lexicon-Based and Support Vector Machine for Assessing the Performance of a Television Program".
- A.K. Jose, N. Bhatia, and S. Krishna, " Twitter Sentiment Analysis ".National Institute of Technology Calicut, 2010.
- P.Lai, "Extracting Strong Sentiment Trend from Twitter", StanfordUniversity,2012.
- Y.Zhou, and Y.Fan, "A Socio linguistic Study of American Slang," Theory and Practice in Language Studies, 3(12), 2209–2213, 2013.
- M.Comesaña,A.P.Soares,M.Perea,A.P.Piñeiro,I.Fraga,andA.Pinheiro,"Author'spersonalcopyComputersinHumanBehaviorERP correlate so masked affective priming with emoticons, "Computers in Human Behavior, 29,13

