

# Comparitive Study on Sentiment Analysis Techniques and User Behavior Prediction on Twitter Data

Aarabhi Babu<sup>1</sup>, Dr.Vince Paul<sup>2</sup>

<sup>1</sup>Mtech student,

Department of Computer Science and Engineering  
,Sahrdaya College Of Engineering and Technology Kodakara,Kerala,India.

[aarabhi.babu@gmail.com](mailto:aarabhi.babu@gmail.com)

<sup>2</sup>Associate Professor

,Department Of Computer Science and Engineering ,  
Sahrdaya College Of Engineering and Technology Kodakara,Kerala,India

[vinceakkara@gmail.com](mailto:vinceakkara@gmail.com)

## Abstract:

People share knowledge, experiences and thoughts with the world by using Social Media like blogs, forums, wikis, review sites, social networks, tweets and so on. This has changed the manner in which people communicate and influence social, political and economic behavior of other people in the Web. This work mainly focus on different sentiment analysis techniques a comparative study of different sentiment analysis techniques and a proposed model that uses sentiment analysis on twitter data and user behavior prediction Proposed System mainly rely on Twitter Data. Performing sentiment analysis on twitter data and predicting the behavior of the tweets and thus the user who post those tweets. By analyzing tweets on certain groups the behavior of those groups can also be identified. Social networking websites are considered as major sources of opinions and views of the public on the prevalent social issues at a given point in time. Websites like the Twitter reflect the public views through its millions of messages posted by its users world wide. By conducting survey these data set required for training data set is being created. and by using these training set twitter commands are analysed and thus the behavior of each tweets are extracted. Based on these tweets if any malicious activities are going on they can be detected and banned.

**Keywords:** Sentiment Analysis, Tweets, Negation Detection Method, parsing

analytics and natural language processing are used to extract and classify;

## 1. Introduction

Social networking websites are considered as major sources of opinions and views of the public on the prevalent social issues at a given point in time. Websites like the Twitter reflect the public views through its millions of messages posted by its users world wide, whenever a controversial issue arises in the society. It is during this time that it is observed significant amount of malicious, violent contents going viral over the internet. In this paper I propose a technique that applies sentiment analysis on data from Twitter and measures the sentiments of posts in order to identify the behaviour of certain groups and users in the twitter. The method goes as follows a training data set is being created manually by conducting a detailed survey and based on this training data set sentiment analysis is performed on the twitter comments then the behaviour of different individuals and groups are extracted and based on the behaviour a graph will be plotted showing the

amount of emotions contained in each group and individuals

### 1.1 ANALYSIS STEPS

The sentiment analysis is a complex process that involves 5 different steps to analyse sentiment data. These steps are:

#### 1.1.1 DATA COLLECTION

The first step of sentiment analysis consists of collecting data from user generated content contained in blogs, forums, social networks. These data are disorganized, expressed in different ways by using different vocabularies, slangs, context of writing etc. Manual analysis is almost impossible. Therefore, text

#### 1.1.2 TEXT PREPARATION

Consists in cleaning the extracted data before analysis. Non-textual contents and contents that are irrelevant for the analysis are identified and eliminated;

#### 1.1.3 SENTIMENT DETECTION

The extracted sentences of the reviews and opinions are examined. Sentences with subjective expressions (opinions, beliefs and views) are retained and sentences with objective communication (facts, factual information) are discarded;

#### 1.1.4 SENTIMENT CLASSIFICATION

In this step, subjective sentences are classified in positive, negative, good, bad; like, dislike, but classification can be made by using multiple points; The main objective of sentiment analysis is to convert unstructured text into meaningful information. When the analysis is finished, the text results are displayed on graphs like pie chart, bar chart and line graphs. Also time can be analysed and can be graphically displayed constructing a sentiment time line with the chosen value over time. So far, the research community has paid little attention to OSN apps specifically

## 2. Sentiment Analysis Techniques

### 2.1 Negation Detection

Negation identification and detecting its scope within a sentence (text) are necessary in finding out the sentiments from

a piece of text. Although negation identification is an important aspect of sentiment analysis, it is yet to be properly addressed.

## 2.2 Bag of Words

It is a technique where each word in a document is represented by a separate variable numeric value (weight) [6]. It is the most widely used technique for sentiment analysis [3, 11]. Das and Chen [3] incorporated negation in their research for extraction of sentiments from stock market message boards. They believed that negation in a sentence reverses the meanings of the sentence. They discussed how words like not, never, no, etc., serve to reverse sentence meaning. They detected negation words in sentences and tag from the sentences with negation markers [3]. In 2002, researchers in [11] adopted the same technique and added the negation word with every word until the first punctuation mark following the negation word. An example that better explains this technique is I do not NOT like NOT this NOT new NOT Nokia NOT model . From the example

above, it can be seen that this technique is not an effective way to find out the negation from a written text as negation may be based on a meaning of words, whereas understanding a scope is very necessary to determine such meanings. Another limitation of this technique is that it is based on the list of words, and lists in any language can never be complete.

## 2.3 VALENCE SHIFTER

Contextual Valence Shifters or modifiers are the words, which change (boost, enhance, diminished.) meanings [8]. Many researchers have transferred their research on sentiment analysis from BOW to Parts of Speech (POS) especially Verbs, Adjectives and Adverbs. The pioneers in giving an understanding that there is a basic polarity associated with every word were in . However, lots of contextual shifters are still needed to change or modify the valance associated with words. Negatives, intensifiers or diminishes are examples of contextual shifter . For example; Negatives: John is clever versus John is not clever. Intensifier: Sam is suspicious about Anna versus Sam is deeply suspicious about Anna Diminishes: I know what to say versus I hardly know what to say. Wigand et al. believed that the effectiveness of the model believed that the effectiveness of the model could be better judged if was evaluated. Kennedy and Inkpen [6] used the same model for Contextual Valence Shifters. They enhanced their model but still kept the scope of any negation term as immediately preceding a term. There is a need for relationship finder to define the scope of negation terms [7]. Other researchers have tried to define the scope by defining lists of verbs, adjectives and adverbs and defining their relationships for sentiment analysis . Lists of positive and negative terms and a set of lists for modifiers was proposed in [8] to define the scope of these modifiers as n- terms before and after positive or negative terms, although this n remained a constant. This technique is better for negation identification in comparison to the BOW technique. However, it also considered the propagation of lists as a limitation. The lists used for this technique may grow with time and can never be complete, as in any language there might be infinite number of words and ways they can be used. Therefore, there is always a need to devise some way for the system to handle words, which are not present in the lists.

## 2.4 ANALYSIS OF NEGATION

For the sentence level sentiment analysis in English language, the basic structure of English sentence and its parts: clauses and phrases are necessary to be understood. These parts further divide sentences into different types of sentences (simple,

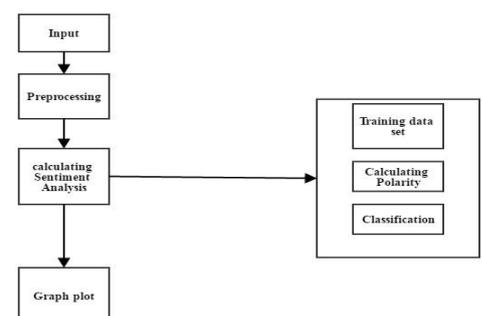
complex and compound). The sentence is made more complex by adding declarative, interrogative, exclamatory and imperative sentences. In order to further complicate the problem as the comparison, contradiction, negation and irony might also be introduced in the sentences. Negation needs to identify its scope, negation can be local (e.g., not good), or it can involve longer-distance dependencies (e.g., does not look very good) or the negation of the subject (e.g., no one thinks that its good). It even changes its roles i.e., instead of negating and can even intensify (e.g., not only good but amazing) [18]. In order to find out the scope of the negation, the sequence of words in the sentence should be identified. On the whole, it is not simply the negation of a word but negation of the sentence . The expression of negation within a sentence varies a lot.

It can be a verb, adverb, suffix or prefix. It might also occur more than once in a sentence and rather than cancelling each other it can give negative meaning, for example; I cannot get no satisfaction [4]. Therefore, the negation analysis has been done using many different ways: Parts of Speech, Bag of Words, and Dependency Tree. However, the best results can be found by combining these approaches. A way to understand the negation by using bag of words approach and latter resolving the scope with the help of dependency tree was proposed in [20]. The following section explains the proposed framework for sentiment analysis and the approach for negation identification and calculation that helps to solve the negation problem in sentiment analysis.

## 3. PROPOSED MODEL

The proposed model is an prediction system that predicts the behavior of user, twitter account holders on twitter .which the behavior of each user and groups are predicted or their character is determined using their tweets only. The tweets of each user is taken till dated and allowed to perform sentiment analysis and based on the result the behavior is predicted so that the wide account holders spreading malicious content and violence promoting groups can be identified and banned. The system works as follows the input of the system contains twitter commends that are collected from twitter tweets and after the data collection these tweets are preprocessed using preprocessing tools. The data will be well cleaned and irrelevant data might be removed after preprocessing the cleaned data then subjected to sentiment analysis in order to calculate the sentiment of the tweets. The sentiment is calculated by using a training data set which was created manually created after conducting survey on the topic , based on the training data set polarity is being calculated for each tweet. Based on the polarity of these tweets they are classified into different categories

**Figure 1** Architectural Diagram



### III. CONCLUSION

By referring different papers of Sentiment analysis different methods and techniques that is been used for performing sentiment analysis was studied methods include using Naïve bayes method, negation detection and calculation, dual sentiment analysis etc. Among this naïve bayes is supervised machine learning technique which can be used for primary classification. where as dual sentiment analysis has considerably better accuracy but its complexity is high so that it is not much requested in online sites like twitter since it takes more processing time. negation detection method parse the given sentence and produce a parse tree and sentiment score is given according to the training data set.

### REFERENCES

- [1] Alessia D'Andrea, Fernando Ferri, Patrizia Grifoni, Tiziana Guzzo Approaches, Tools and Applications for Sentiment Analysis Implementation ", International Journal of Computer Applications (0975 8887) Volume 125 No.3, September, 2015.
- [2] Rui Xia, Feng Xu, Zong, and Qianmu Li. Dual Sentiment Analysis: Considering Two Sides of One Review ". IEEE TRANSACTIONS 2015.
- [3] Efthymios Kouloumpis, Theresa Wilson, Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG", Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011
- [4] Ashish Shukla, Rahul Mishra. Sentiment Classification and Analysis Using Modified Kmeans and Naive Bayes Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, 2015.
- [5] Shilpa P, Madhu Kumar S D. Feature Oriented Sentiment Analysis in Social Networking Sites to Track Malicious Campaigners", 2015 IEEE Recent Advances in Intelligent Computational Systems (RAICS) | 10-12 December, 2015.
- [6] Bo Pang and Lillian Lee . Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, 2008.
- [7] M. Gjoka, M. Sirivianos, A. Markopoulou, and X. Yang. Poking facebook: characterization of osn applications.", In Proceedings of the rst workshop on Online social networks, WOSN, 2008.
- [8] Zohreh Madhoushi, Abdul Razak Hamdan, Suhaila Zainudin. Sentiment Analysis Techniques in Recent Works", In Proceedings of the rst workshop on Online social networks, 2015.
- [9] Amna Asmi, Tanko Ishaya. Negation Identification and Calculation in Sentiment Analysis.", The Second International Conference on Advances in Information Mining and Management, 2012.
- [10] Yahya Eru Cakra, Bayu Distiawan Trisedya. Stock Price Prediction using Linear Regression based on Sentiment Analysis.", ICACIS, 2015
- [11] M Karamibekr and Ghorbani A. Ali. Verb Oriented Sentiment Classification. In IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology (WI-IAT), pages 327 -331, May 2012.
- [12] Tom M. Mitchell. Machine Learning. McGrawHill Inc., New York, NY, USA, First edition, 1997.

- [13] Elizabeth D. Liddy. Natural Language Processing. In Encyclopedia of Library and Information Science, First edition, 2001.