

# A Survey on Analysis of Twitter Opinion Mining Using Sentiment Analysis

Anusha K S<sup>1</sup>, Radhika A D<sup>2</sup>

<sup>1</sup>M Tech, CSE Dept. of Computer Science and Engineering VVCE, Mysuru

<sup>2</sup>Assistant Professor, CSE Dept. of Computer Science and Engineering VVCE, Mysuru

\*\*\*

**Abstract** - In recent years, there is a rapid growth in online communication. There are many social networking sites and related mobile applications, and some more are still emerging. Huge amount of data is generated by these sites everyday and this data can be used as a source for various analysis purposes. Twitter is one of the most popular networking sites with millions of users. There are users with different views and varieties of reviews in the form of tweets are generated by them. Nowadays Opinion Mining has become an emerging topic of research due to lot of opinionated data available on Blogs & social networking sites. Tracking different types of opinions & summarizing them can provide valuable insight to different types of opinions to users who use Social networking sites to get reviews about any product, service or any topic. Analysis of opinions & its classification on the basis of polarity (positive, negative, neutral) is a challenging task. Lot of work has been done on sentiment analysis of twitter data and lot needs to be done.

In this paper we discuss the levels, approaches of sentiment analysis, sentiment analysis of twitter data, existing tools available for sentiment analysis and the steps involved for same. Two approaches are discussed with an example which works on machine learning and lexicon based respectively.

**Keywords**—*Twitter data, opinion mining, sentiment analysis.*

## I. INTRODUCTION

Sentiment analysis technique is an effective means of discovering public opinions. Various companies often use online or paper based surveys to collect customer comments. Due to the emergence of social networking sites and applications, people tend to comment on their facebook or tweet profile. Therefore the paper based approach is not an efficient approach. Only a very small customer base can be reached and there is no guarantee that their answers in the survey are honest or not. Here social media comes into play. Facebook, Twitter and all other social media sites are full of people's opinions about products/services they use, comments about popular personalities and much more [1].

Nasukawa & Yi first introduced the term Sentiment Analysis & Opinion Mining in the year 2003. Opinion mining tools will help in providing the opinion about the product. Sentiment analysis on tweet data involves data collection, extraction, classification, understanding and providing the opinion that are expressed in various tweets [2].

## Opinion Mining and Sentiment analysis is done on three levels [3]

- Document Level: Analysis is done on the whole document and then expresses whether the document is positive or negative sentiment.
- Sentence Level: It is related to find sentiment polarity from short sentences. Sentence level is merely close to subjectivity classification.
- Entity /Aspect Level: sentiment analysis performs augmented analysis. The aim is to find sentiment on entities or aspects.

## Two approaches of Sentiment Analysis

- **Supervised approaches or machine learning method:**

Machine learning is one of the most prominent techniques gaining researchers interest [10] due to its adaptability and accuracy. This method comprises of three stages: (i) Data collection (ii) Pre-processing and (iii) Training data Classification [9].

- **Unsupervised (or lexicon-based):**

Lexical analysis estimates the sentiment from the semantic orientation of words [8] or phrases that occur in a text. In this approach a dictionary containing positive and negative words that are matched with the words containing in tweet. However, these techniques totally depend on lexical resources [6] which are concerned with mapping words [7] to a categorical (positive, negative, neutral) or numerical sentiment score. In this method the unigrams, which are found in the lexicon [9] are assigned a polarity score.

## Sentiment Analysis of twitter data

Twitter is popular online social networking service launched in March 2006. It enables users to send and read tweets with about 140 characters length. Currently Twitter acts as opinionated Data Bank with large amount of data available used for sentiment analysis. Twitter is very convenient for research because there are very large numbers of messages, many of which are publicly available, and obtaining them is technically simple compared to scarping blogs from the web.

Twitter data is collected for analysis using Twitter API.

## II. Tools Available for sentiment analysis

S. No	Tools	Applications
1	NLTK	NLTK toolkit is widely used nowadays for sentiment analysis task. Main features of NLTK used in Sentiment analysis process are Tokenization, Stop Word removal, Stemming and tagging. This tool is written in Python language and can be downloaded from <a href="http://www.nltk.org">www.nltk.org</a> .
2	GATE	General Architecture for Text Engineering (GATE) is information Extraction System, Stemming and Part of speech tagger. This tool is written in Java language. <a href="https://gate.ac.uk/">https://gate.ac.uk/</a>
3	Red Opal	This tool is widely used for users who want to buy any products based on different features. Users can search for any product depending upon the feature selected and can get reviews related to their search.
4	Opinion finder	Opinion Finder is used for analysis of different Subjective sentences related to any topic & classification of sentences is done based on their polarity. It's written in Java and is platform Independent tool.

Table-1: Sentiment Analysis Tools

## III. TWITTER SENTIMENT ANALYSIS PROCEDURE

The framework used for this analysis is depicted in below figure. Different processing steps had their own important role. We discussed about all steps below.

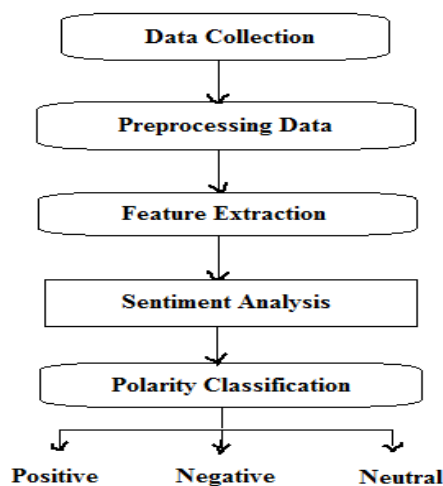


Figure-1 Sentiment analysis Framework

### A. Data Collection:

Collection of data is an important part of Sentiment Analysis. Various data Sources like Blogs, Review Sites, Online Posts & Micro Blogging like Twitter, Facebook are used for Data Collection.

### B. Data Preprocessing:

Now before Sentiment Analysis we need to process the collected data using the following steps of data processing-

1) Stemming- In this process the postfix from each words like "ing", "tion" etc are removed.

2) Tokenization- This process is very important for Data pre processing as it includes several sub steps like "Removal of Extra spaces", "Emoticons (☺,☹) used replaced with their actual meaning like Happy, Sad by using Emoticon data set available on Internet", "Abbreviations like OMG, WTF are replaced by their actual meanings", "Pragmatics handling like hapyyyyyyy as happy, guddddd as good etc."

3) Stop Word Removal- In this, stop words which are not of any use in analysis like Prepositions (a, an) and Conjunctions (and, between) used are removed.

### C. Feature Extraction:

Feature extraction specifies the type of features used for opinion Mining [6]. There are different types of features used like-

1) Term Frequency- Frequency of any term in a document carries weight age. [6]

2) Term Co-occurrence- Repeatedly occurrence of a word like Unigram, Bigram or n-gram etc. [6]

3) Part of Speech- For each tweet we have features for counts of the number of Verbs, adjectives, nouns. [7]

### D. Sentiment Analysis & Polarity Classification:

Emotions, opinions and sentiments play an important role in all human life. Mining such opinions termed as sentiment analysis [10]. Performing task of Sentiment analysis & polarity classification is a challenging task. SentiWord net is a standard dictionary used by most researchers today for sentiment analysis. Task of Polarity classification mean the reviews collected are classified depending upon the emotions expressed as Positive, Negative and Neutral.

#### IV. APPROACHES USED FOR SENTIMENT ANALYSIS ON TWITTER

##### A. SENTIMENT ANALYSIS ON TWITTER USING SUPERVISED APPROACH (MACHINE LEARNING)

This approach extracts the data from SNS services which is done using Streaming API of twitter. The extracted tweets are loaded into hadoop and it is been preprocessed using map reduce. This task is followed by classification which uses NLP and machine learning techniques. The classification used here is uni-word naïve bayes' classification.

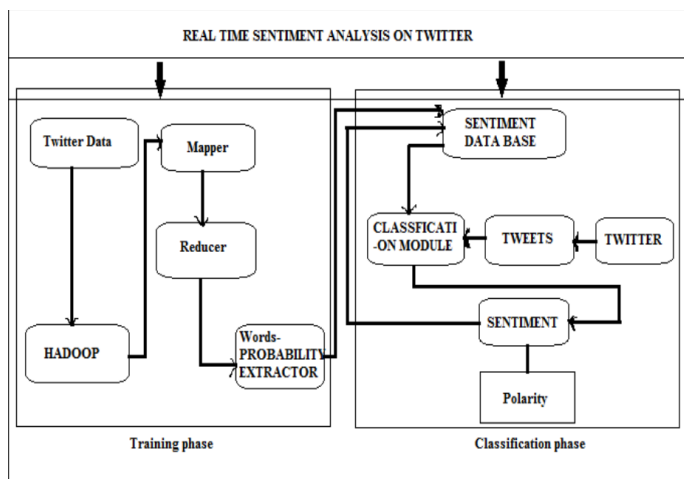


FIGURE 2: real time sentiment analysis on twitter

Consider the number of all positive tweets, positive words and negative words from our training phase. Then calculate the probability of a tweet being positive.

$$P(C) = \text{No. of Positive Tweets} / \text{Total Number of Tweets} \quad (1)$$

For each word in each tweet that is being streamed is checked for the probability of it being used given that it is positive.

$$P(D/C) = \text{Positive score of the words} / \text{Total number of Positive Words} \quad (2)$$

Then checked the word itself being used irrespective of whether or not it is positive.

$$P(D) = \text{Positive score of the word} + \text{Negative score of the Word} / \text{Total Number of words} \quad (3)$$

In-order to check the probability of word being positive given that is used in a tweet which is given as follows:

$$P(C/D) = P(C) * P(D/C) / P(D) \quad (4)$$

The probability of a word is then passed to the Sentiment function which then classifies, if the probability of the word is greater than 0.6 then it is positive, as neutral if the probability is between 0.4 and 0.6 and negative if it is lesser than 0.

##### B. SENTIMENT ANALYSIS ON TWITTER USING UNSUPERVISED APPROACH (LEXICON METHOD)

The data is collected from the twitter API and that data is pre-processed to eliminate all unwanted information and to replace the emoticons. Here lexical method is used for classification and work on dictionary-based approach. The dictionary-based approach depends on finding words from tweets, and then matches the word with the dictionary. If there is a positive match, the positive score is shown or the word is tagged as positive. If it is negative word then the negative score is incremented or the word is tagged as negative. Otherwise tag neutral word.

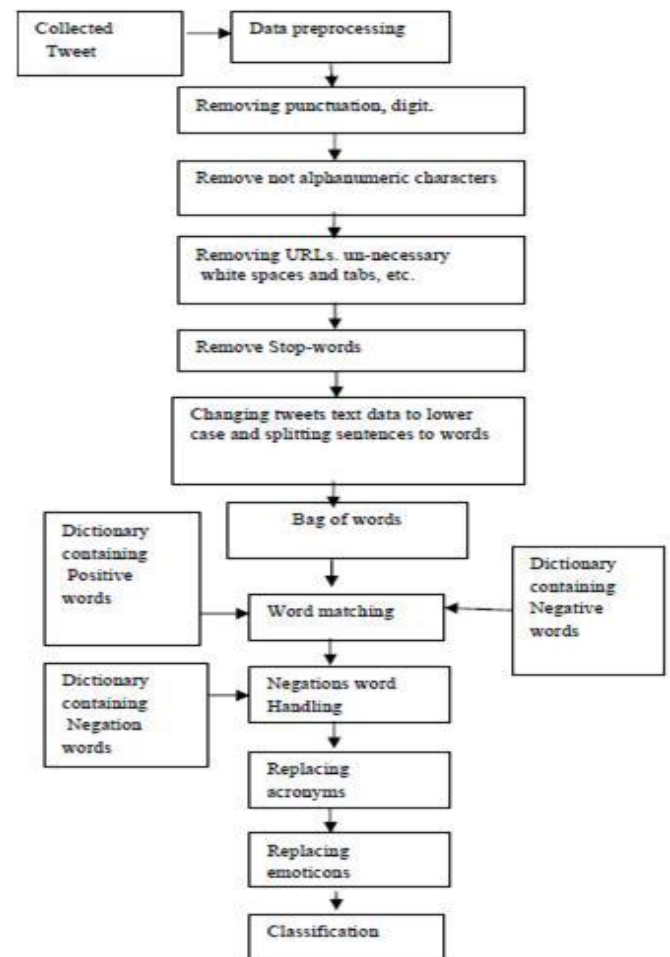


FIGURE 3: working of a lexical technique

Let a sentence  $s$  contains a set of entities  $\{e_1, e_2, \dots, e_r\}$  and a subset of their aspects  $\{a_1, \dots, a_m\}$  from a set of opinion holders [2]  $\{h_1, h_2, \dots, h_p\}$  at some particular time point and a set of sentiment words or phrases  $\{sw_1, \dots, sw_n\}$  with their sentiment scores. The sentiment orientation for each aspect  $a_i$  in  $s$  is determined by the following aggregation function:

$$score(a_i, s) = \sum_{sw_j \in S} \frac{sw_j \cdot so}{dist(sw_j, a_i)}$$

Where  $\text{dist}(\text{swj}, a_i)$  is the distance between aspect  $a_i$  and sentiment word  $\text{swj}$  in  $s$ .  $\text{swj}$ .so is the sentiment score of  $\text{swj}$ . If the final score is positive, then the opinion on aspect  $a_i$  in  $s$  is positive. If the final score is negative, then the sentiment on the aspect is negative. It is neutral otherwise.

## CONCLUSION

In this paper, we have done a short survey on sentiment analysis and opinion mining. We have discussed about three major levels of sentiment analysis, two approaches of sentiment analysis and sentiment analysis of twitter data. Further, we studied and listed some of the tools available for sentiment analysis and the general procedure for sentiment analysis. By analyzing the examples of both approaches we understood the two approaches in detail.

## REFERENCES

- [1] Syed Akib Anwar Hridoy, M.Tahmid Ekram, Mohammad Samiul Islam, Faysal Ahmed and Rashedur M. Rahman "Localized twitter opinion mining using sentiment analysis".
- [2] Roshan Fornandes, Dr. Rio D'Souza "Analysis of product twitter data through opinion mining"©2016 IEEE.
- [3] M. Trupthi, Suresh Pabboju, G. Narasimha "Sentiment analysis on twitter using streaming API" 2017 IEEE &th International Advance Computing Conference.
- [4] Perna Mishra, Dr. Ranjana Rajnish, Dr. Pankaj Kumar "Sentiment analysis of twitter data: Case study on digital india" 2016(InCITe).
- [5] Paramita Ray, Amlan Chakrabarti "Twitter sentiment analysis for product reviews using Lexicon Method" 2017(ICDMAI).
- [6] A Kowcika and Aditi Guptha "sentiment Analysis for social media" ,International journal of advanced research in computer science and software engineering,216-221,Volume 3,Issue 7, july 2013.
- [7] G. Vinodini and RM.Chandrashekar, "sentiment analysis and opinion mining: A survey", International journal of advanced research in computer science and software engineering,283-294, Volume 2,Issue 6, june 2012.
- [8] Cataldo Musto, Giovanni Semeraro, Marco Polignano, "A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts", Department of Computer Science, University of Bari Aldo Moro, Italy.
- [9] James Spencer and Gulden Uchyigit, Sentimentor: Sentiment Analysis of Twitter Data. School of Computing, Engineering and Mathematics. University of Brighton.
- [10] Anna Jurek, Maurice D. Mulvenna and Yaxin Bi, Improved lexiconbased sentiment analysis for social media analytics Science direct, Published: 9 December 2015.
- [11] Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau, "Sentiment Analysis of Twitter Data", Columbia University, Newyork.
- [12] Sang-Hyun Cho and Hang-Bong Kang, "Text Sentiment Classification for SNS-based Marketing Using Domain Sentiment Dictionary", IEEE International Conference on Conference on consumer Electronics(ICCE), p.717-718, 2012.
- [13] Patricia L V Ribeiro, Li Weigang and Tiancheng Li "A Unified Approach for Domain-Specific Tweet Sentiment Analysis", FUSION, 2015.
- [14] Tiara, Mira Kania Sabariah, Veronikha Effendy, "Sentiment Analysis on Twitter Using the Combination of Lexicon-Based and Support Vector Machine for Assessing the Performance of a Television Program", 3<sup>rd</sup> International Conference on Information and Communication Technology (ICoICT), 2015.
- [15] Asmita Dhokrat, Sunil Khillare, C. Namrata Mahender, "Review on Techniques and Tools used for Opinion Mining," IJCAT, 2015.