Impact Factor Value 4.046 **e-ISSN: 2456-3463**
National Conference on "Recent Advances in Engineering and Technology" SAMMANTRANA 19
Organized by Government College of Engineering, Nagpur
International Journal of Innovations in Engineering and Science, Vol 4 No.8, 2019
www.ijies.net

A Review of Opinion Mining On Twitter Using Lexicon Based Approach

AshleshaWadurkar[1], Akash Bhad[2], Bhavana Rajput[3], Durgesh Pandey[4], Dr. Sachin V. Solanki[5]

[1,2,3,4,5] *Dept. of Information Technology KDKCE, Nagpur, India, 440009*

Abstract *–Social networking sites are becoming a large source of raw data. Millions of people nowadays share their views on social media and blogging (micro blogging) because its short and simple to use. Twitter is one of the micro blogging sites where users posts their views in form of tweets. The rapid increase in such data on social media generates a need of mining such data to get valuable results. Mining such data can provide very useful insights for different applications. This paper presents a review of opinion mining techniques over twitter data. The data present on twitter can be unstructured and may need pre-processing. Opinion mining addresses such need by extracting raw data, performing different pre-processing techniques and detecting the opinions/emotion/ sentiments of textual data. This mined i.e. sentiment of people over a particular topic can be helpful in various fields such as social, medical and industrial applications. This paper presents a survey about the opinion mining, its techniques, different concepts this field, problems and its solution.*

Keywords- Opinion mining, Twitter, Micro blogging, Social media, Pre-processing, Sentiments

INTRODUCTION

Opinion mining is a field of natural language processing for evaluating the mood of public about a particular thing. Opinion mining which is also known as sentiment analysis involves developing a system to collect and categorize opinions about anything. Opinion mining can be useful in many ways. It can help businessmen to evaluate product's an ad campaign, what's consumer's priority is, their likes and dislikes etc. In detail opinion mining is a method of extracting the sentiments and categorizing those sentiments into positive, negative and neutral.

Internet and Online communities are huge sources of information. A large number of datasets can be acquired from various e- commerce websites and social networking sites such as online product reviews and tweets, posts and comments and blogs. These can be captures to analyse the sentiments of the author/writer. Extracting opinions and sentiments from online texts requires different levels of pre-processing, classification of text and methods for text mining approaches[1].

Social networking is a growing computing for the analysis and modelling of various social activities on different environment[3] Enormous amount of data is available on social media due to the increasing uses of it. This Rapid increase in data over social media can mined and its results can be useful to find out many future insights. Our social media is mostly in unstructured form[1]. This makes it difficult to analyse and acquire valuable insights from this data. Such data can then mind, and user sentiments / opinions can be extracted.

Textual information present on twitter can be majorly classified into one of the given categories: fact data and sentiment data [9].Fact data are basically the objective terminologies concerning different entities, things or events. On the other hand, the sentiment data are the subjective terms, that defines a person's opinions or belief or sentiments for a particular entity, thing, events or product. Sentiment analysis has three main aspects i.e. the subject sentiment holder, sentiment itself and object i.e. the topic about which the subject has shared their opinion[3].

Twitter is a very popular micro blogging site that has rapidly growing users. Users express their views and daily status on twitter in the form of tweets. Tweets are short messages that describes what's on user's mind.

Tweets can be the user's opinion towards product, events, things, services and other twitter users in which they are interested. There are many cases in which user's opinion can be useful for many real world application such as product /service reviews on restaurants, electronics, etc[11].The main objective of this resea5rch is to analyze the lexicon based method for opinion mining over twitter data.

Opinion mining on twitter is useful in various ways and fields such as:

- Movies: Is the movie review positive or negative?
- Public sentiment: Is despair increasing?
- Political: What do people think about a particular candidate in politics and any political issue raised?
- Products: What do people the customers think about any new product in market. E.g.: Phone

LITERATURE SURVEY

- **Background Study:**

1. Opinion mining: Opinion mining refers to the area of NLP, text mining, computational linguistics, which involves the computational study of sentiments, opinions and emotions of people expressed in the form of text.[7] the opinion / sentiment are the attitudes of people depending upon emotions, delete and common understanding. OM has many applications in various fields such as social, medical specific application domain of opinion mining includes accounting, law, entertainment, education, technology, politics and industrial application and marketing. [7]

2. Social media: It is defined as social media as group of internet leased applications that create on the ideological and technological foundation of web 2.0 which is allowed to build an exchange of user generated content. There has been a rapid increase in data on social media due to the increased use of Technology. Most of the data that exist over social media is in unstructured form. Such unstructured data over social media is approximately 80% of the data over world. The nature of the data over social media is unstructured creates the need for mining.

3. Twitter: Twitter is a very popular micro blogging social networking site which allows its users to express their views and opinions on various topics. Fruits which are small text information 80 characters[7] twitter is keypad phone networking sites from where the general opinion of general public related to their day to day life as well as opinions about specific topic can be extracted.

4. Twitter opinion mining: the opinion mining over Twitter can be done by using the comments/ tweets for providing useful indicators for many different purposes. The tweets can be categorised into positive, negative and neutral and can provide insights for various applications. Sentiment analysis / opinion mining is a natural language processing technique which quantifies the opinion of expressed tweets over Twitter [7]. Opinion mining refers to a method of extracting polarity and subjectivity from words, text or phrases over Twitter [7].

- **Opinion mining levels:**

Sentiment analysis and opinion mining can be applied in four levels which detects the positive, negative and neutral sentiment depending on the level.

Level 1 is sentence level. This level detects the positive, negative, neutral sentiments/opinions of a sentence[1]. For a given sentence it categorizes the sentence into positive, negative, neutral and finds out the polarity value. Level 2 is the document level which applies the analysis process to the whole document in finds its sentiment considering the whole document as a single unit or one entity. Document
is categorized either positive or negative. Level 3 is aspect level. This level is used when there are attributes of an entity. Each attribute has their own different sentiments. For example:

- **Existing work:**

Khaled Ahmed[1] in his Review paper "Sentiment Analysis over social network: anoverview". The rapid increase of data in unstructured form on social media creates need a mining to get valuable data. In this paper the Author has describe the survey of sentiment analysis addressing the Application of sentiment analysis in different domains like social, medical and industrial. It also describes the problem and its solution in different areas available APIs, tools used and presenting a list of open challenges in this area.

Mithali Desai in her paper **"**Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey**"** [3] focuses in unstructured data on Twitter. Secondly describe various techniques to carryout sentiment analysis on Twitter data in detail. Moreover, he presents the parametric comparison of the discussed techniques based on our identified parameters. The author discussed the various technique of sentiment analysis.

M S Neethu in her paper "Sentiment Analysis in twitter using machine learning"[5] attempt to analyze the twitter posts about electronic products like mobiles, laptops etc using Machine Learning approach. By doing

sentiment analysis in a particular domain, it is possible to identify the effect of domain information in sentiment classification. We represent a new feature vector for classifying the tweets as positive, negative and extract people'sviews about products.

Alisa Sarlan in paper "Twitter Sentiment Analysis"[7] reports on the design of a sentiment analysis, extracting a vast number of tweets. Prototyping is used in this development. Resultsdifferentiate consumers perspective via tweets into positive and negative,which is represented in a pie chart and html page. However, the program has planned to establish on a web application system, but due to limitation of Django which can be worked on a Linux server or LAMP, for further this approach need to be done.

Author Yusuf Arslan in his paper "Real time Sentiment Analysis On twitter" [12] has used twitter specific dictionaries constructed manually i.e. positive and negative dictionaries containing positive and negative sentiments including emotions, slangs, abbreviation and misspelled words. They have used a reverse strategy by matching unigrams and bigrams in the tweet pool/twitter data set. They have also used SentiWordNet to make a comparison of their twitter specific dictionary and SentiWordNet vocabulary.

METHOLOGY

A. Twitter

Twitter is a social networking site, where users post real time reactions to and opinions about "everything". Tweets were originally restricted to 140 characters, but on later, this limit was doubled for all languages except Chinese,Japanese,and Korean. Registered consumerscan post, like, and retweet tweets, but unregistered consumers can only read them. Consumer access Twitter through its website interface, through Short Message Service (SMS) or its mobile-device application software. Over the last year, Twitter has made a number of changes, small and big, to drive user engagement and improve the overall on boarding and experience of the platform.

B. Twitter Data Collection Methods

The three possible ways to gather Tweets for research are as follows [13]:

• Data repositories like UCI, Friendster, Kdnuggets, and SNAP.

• APIs: Twitter gives two types of APIs such as search API and stream API. Search API is used for collect Twitter data on the basis of hash tags and stream API is used for stream real time data from Twitter.

• Automated tools differentiated into premium tools such as Radian6,[14]Sysmos, Lithium and non-premium tools like Keyhole, Topsy, Tagboard and Social Mention.

C. Data Pre-processing

Mining of Twitter data is a very tedious task. The collected data is raw data. So as to apply classifier, it is necessary to pre-process or clean the raw data. The pre-processing task involves consistent casing, removal of hashtags and other Twitter notations (@, RT), emoticons, URLs, stop words, decompression of slang words and compression of elongated words. The following steps show the pre-processing procedure:

- Remove the Twitter informationlike hashtags (#), retweets (RT), and account Id (@).
- Remove the URLs, hyperlinks and emoticon. It is essential toeliminate on letter data and symbols as we are dealing with only text data.
- Remove the stop words such as is, am, are etc. The stop words do not focus on any emotions, it is necessary to remove them to reduce.
- Reduce the elongated words such as cooool into cool.
- Decompress the slag words such as gr8, f9. Generally, slang words are adjectives or nouns and they have extreme level of sentiments. So it is necessary to clean them.

D. Feature Extraction:

After the data is pre-processed, it has various discrete properties.[3] In the feature extraction methods, the different aspects such as nouns, adjectives, verbs are extracted and later these aspects are recognized as positive or negative to notice the polarity of the whole sentence. Followings are the widely used Feature Extraction methods.

- Parts Of Speech (POS):It refers to the finding of nouns, verbs, adjectives etc. as they are significantly important for analysis.
- Negative Phrases: The existence of negative words can change the meaning of the opinion. So, it is essential to consider negative words.

E. Approach for Opinion Mining: Lexicon Based Approach:

Lexicon based approach is a chief method for Opinion Mining. It is one of the two main processes of Sentiment analysis. It involves analyzing and calculating the sentiments from the semantic orientation of sentences

and words in a text unit[10]. This method mainly revolves around the concept of creating lexicons out of the sentences and phrases. It also incorporates methods for dictionary creation one which includes all the positive, negative and neutral words having a sentiment value. The basic strategy is to compare these lexicons with the sentiment words in the dictionary and find their sentiment value. Lexicon based method is more precise, easy to understand and easily implementable[8] but the problem is that it requires high level of human interaction in the text analysis approach. This approach mainly deals with phrases, expressions or textual contents that are generally found in chats, product reviews, posts on social media etc.

Dictionary Based: This method/approach includes creation of sentiment dictionaryusing positive, negative and neutral word having some values assigned to them which are their sentiment values. These sentiment dictionary can be created manually by developer or some predefined dictionary such as WordNet can be used.[8]

Corpus Based Approach: This method/approach typically incorporates corpus data inorder to explore a theory or hypothesis. Corpus is nothing but a large dataset which can be analyzed.[8]

FLOW

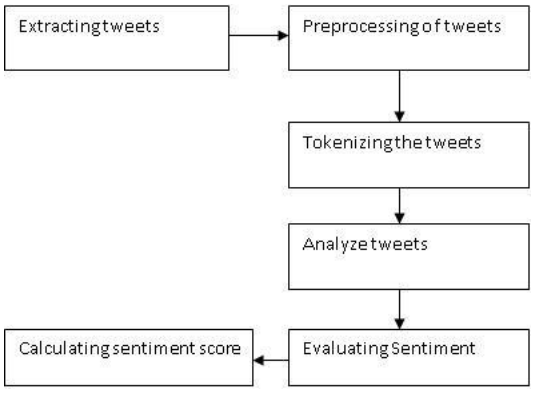

Figure: 4.1 Flow of the system

CHALLENGES

- Difficulty of opinion mining with inappropriate English.
- Applying opinion mining on short forms/ abbreviation.
- Identifying sarcastic sentence.
- Identifying fake comments.
- Appropriately resolving co reference in the usage of nouns and pronouns.
- Diversity of language.

CONCLUSION

This paper presents the overall procedure to perform Opinion Mining process to categorize unstructured data of Twitter into positive or negative categories.Secondly, we have discussed the lexicon based technique to perform opinion mining on Twitter data. It has been found that the technique applied for opinion mining is language specific. Hence, in future the opinion mining can be implemented for diverse languages as well. Language variety in social media data is a key issue which is required to be eliminated in future.

REFERENCES

[1] Khaled Ahmed, Neamat El Tazi," Sentiment Analysis Over Social Networks: An Overview", 2015 IEEE International Conference on Systems, Man, and Cybernetics.

[2] Rushlene Kaur Bakshi, Ravneet Kaur, "Opinion mining and Sentiment Analysis", 2016 International Conference on Computing for Sustainable Global Development (INDIACom).

[3] MitaliDesai,Mayuri A. Meheta," Techniques for Sentiment Analysis of Twitter Data: A Comprehensive Survey "International Conference on Computing, Communication andAutomation (ICCCA2016) ISBN: 978-1-5090-1666-2/16/$31.00 ©2016 IEEE 149

[4] BalakrishnanGokulakrishnan ,PavalanathanPriyanthan, ThiruchittampalamRagavan,"Opinion Mining and Sentiment Analysis on a Twitter Data Stream ", The International Conference on Advances in ICT for Emerging Regions - ICTer 2012 : 182-188.

[5] Neethu M S, Rajasree R," Sentiment Analysis in Twitter using Machine Learning Techniques" IEEE – 31661, 4th ICCCNT 2013 July 4 - 6, 2013, Tiruchengode, India.

[6] Neha Raghuvanshi, Prof J. M. Patil," A Brief Review on Sentiment Analysis" InternationalConference on Electrical, Electronics, and Optimization Techniques (ICEEOT) – 2016 978-1-4673-9939-5/16/$31.00 ©2016 IEEE.

[7] Aliza Sarlan, ChayanitNadam, ShuibBasri," Twitter Sentiment Analysis" 2014International Conference on Information Technology and Multimedia (ICIMU), November 18 – 20, 2014, Putrajaya, Malaysia 978-1-4799-5423-0/14/$31.00 ©2014 IEEE 212

[8] Shantanu Mandal ,SumitGupta,"A Lexicon-Based Text Classification Model to Analysis and Predict Sentiments from Online Reviews",IEEE, December, 2017.

[9] Bing Liu, N. Indurkhya and F. J. Damerau, Handbook of Natural Language Processing, Second Edition, 2010, pp. 1-3860-68.

[10] Deptii D. Chaudhari et al., "Feature Based Approach for Review Mining Using Appraisal Words", International Conference on Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), IEEE 2013.

[11] B. Pang and L. Lee, -Opinion Mining and Sentiment Analysis. In Foundations and Trends in Information Retrieval, vol. 2, pp. 1-135,2008.

[12] *Yusuf Arslan, AysenurBirturk, BekjanDjumabaev, DilekKuc,"Real-Time Lexicon-Based Sentiment Analysis Experiments On Twitter With A Mild (More Information, Less Data)Approach" 2017 IEEE International Conference on Big Data (BIGDATA)*

[13] "Three Cool and Inexpensive Tools to Track Twitter Hashtags", June 11, 2013. [Online]. Available http://dannybrown.me/2013/06/11/three-cool-toolstwitterhashtags/ [Accessed: 19-Oct-2015].

[14] B. Gokulakrishnan, P. Plavnathan, R. Thiruchittampalam, A. Perera and N. Prasath, "Opinion Mining and Sentiment Analysis on aTwitter Data Stream", in Int. Conf. on Advances in ICT for Engineering Regions, 2012, pp. 182-188.