Impact Factor (SJIF): 4.542

International Journal of Advance Research in Engineering, Science & Technology

e-ISSN: 2393-9877, p-ISSN: 2394-2444
Volume 4, Issue 3, March-2017

A Review on Web Sentiment Analysis for Scoring Positive or Negative Words using Social Networking Sites

Parag Kadu[1], V.S.Sakharkar[2], S.V.Baghel[3],R.R.Karwa[4],G.B.Saboo[5]

[1] *CSE Department, P.R.M.I.T & R College*
[2] *CSE Department, P.R.M.I.T & R College*
[3] *CSE Department, P.R.M.I.T & R College*
[4] *CSE Department, P.R.M.I.T & R College*
[5] *CSE Department, P.R.M.I.T & R College*

Abstract — ***As people are free to say their opinions on anything using various social networking sites like Twitter, Facebook , Discussion forums, and blogs. Particularly Micro blogging and text messaging have emerged and become dominated tool over the web. Post are often used to share opinions and sentiments about the surrounding world. The availability of social content generated on social sites creates new opportunities to study public opinion about the entity. We took comments as a data for sentiment classification. The Sentiment analysis is done on a per comment basis. The words in each comment are compared after looking at these words, the algorithm then judges whether the text in the post is positive or negative based on the likelihood for each possibility.***

Keywords- Sentiment Analysis, Natural Language Processing, Artificial Intelligence

I. INTRODUCTION

Nowadays, Social media is becoming more and more popular since mobile devices can access social network easily from anywhere. Therefore, Social media is becoming an important topic for research in many fields. As number of people using social network are growing day by day, to communicate with their peers so that they can share their personal feeling everyday and views are created on large scale. In this scientific era, internet provides huge volume of information. Most of the people share their opinions over internet by using social networking sites in form of textual data. These textual data are publically available over internet & has a great impact in building opinions about a particular entity, object or political activities among the users of social media. Shared information is generally in form of reviews, articles, posts, news etc [5].

Fig 1: Various Social networking sites

Social media user may find others opinion by collecting & analyzing their reviews. Social media monitoring has been growing day by day so analyzing of social data plays an important role in knowing customer behavior. So we are analyzing Social data such as facebook ,Twitter Tweets using sentiment analysis which checks the attitude of User and their opinion on particular post, product, movie etc [9].

Over the recent years, an emerging interest has been occurred in supporting social media analysis for advertising, opinion analysis and understanding community cohesion. Social media data adapts to many of the classifications attributed for "big-data" – i.e. volume, velocity and variety [13]. Analysis of Social media needs to be undertaken over large volumes of data in an efficient and timely manner. Analyzing the media content has been centralized in social sciences, due to the key role that the social media plays in modeling public opinion. This type of analysis typically on the preliminary coding of the text being examined, a step that involves reading and annotating the text and that limits the sizes of the data that can be analyzed.

In the decision making process each and every piece of information are very important [9].

Fig 2: Connected people by networking sites

After arriving internet world user doesn't bother about other opinions from individuals newspaper, surveys, opinion pools, consultants because web analytics introduce new system called opinion mining, which is find out the opinions and experience of other people over the internet using digital social media network like Facebook , reviews, forums, blogs, Twitter, micro-blogs, etc., Indeed, according to surveys about 6 in 10 (60%) online shoppers say user generated customer product reviews have a significant or good impact on their buying behavior [1][2].

Also Data from the 2011 Social Shopping Study indicates that 50% of consumers spend 75% or more of their total shopping time conducting online product research, with 15% spending 90% or more of their shopping time in this manner. Another surveys by Deloitte Consumer Products Group found that almost two-thirds (62%) of consumers read consumer written product reviews online. In fact, a recent study by Deloitte found that "82% of purchase decisions have been directly influenced by reviews". The objective is to throw lime light on determine the sentiment of the text, whether it is positive or negative, which is extended to strength of polarity. With the explosion of Web 2.0 platforms such as blogs, discussion forums, peer-to-peer networks, and various other types of social media [15].

Consumers have at their disposal a soapbox of unprecedented reach and power by which to share their brand experiences and opinions, positive or negative, regarding any product or service. As major companies are increasingly coming to realize, these consumer voices can wield enormous influence in shaping the opinions of other consumer and, ultimately, their brand loyalties, their purchase decisions, and their own brand advocacy. Companies can respond to the consumer insights they generate through social media monitoring and analysis by modifying their marketing message, brand positing, product development, and other activities accordingly [3][4] .

II. SYSTEM ANALYSIS

2.1. Existing System

We evaluate Friend book system on large scale experiments using social media data. Recently analysis is worked for few Twitter tweets but here we are to doing work for user behavior analysis using Facebook Comments or any social media data reviews, user likes, interests , Ecosystems for getting raw data from the

Facebook or any social media sites we are using online streaming tool using Apache Flume [6]. Using Flume tool only we configure everything that we want to get data from the Facebook [7]. For analysis we want to set the configuration and also want to define what information that we want to get form Facebook all these will be saved [8] in our prescribed format. In this system, a method to calculate Analysis of reviews or comments given by the customers or user is proposed and implemented in Java

2.2. Proposed System

The main goal of this system is to perform sentiment analysis on the comments collected from various social networking sites. In this system, searching the information based on category and keywords from the database is performed. Searching keywords in database is one of the hardest tasks because of the diversity of the language and the slangs used on the internet. In this system, the first step involves collection of comments from social networking sites and making it as a data set, the second step is preprocessing of the related .

In the third step, applying different mining techniques to derive usefulness about stored information. Different mining approaches are classification, clustering, statistical analysis, natural language processing etc [11]. In the fourth step, similar data is identified and analyzed, then by using a web application, the final results are shown.

Fig 3: Database of Various Social networking sites

We are using positive and negative words dictionary to identify positive and negative words [9] [10]. Stop word dictionary is used to identify and remove stop words from the reviewed product [11]. The focus of our project is to assign the polarity to each comment i.e. whether the author express positive or negative opinion [12].

III. IMPLEMENTATION AND RESULT

3.1 Data Acquisition

In this data acquisition, data are gathered from different relevant sources such as web crawling, twitter tweets, online review, newsfeeds, document scanning etc [6].

3.2 Preprocessing

It is used to remove noisy, inconsistent and incomplete data. For doing the classification, Text preprocessing and feature extraction is a preliminary phase [4].

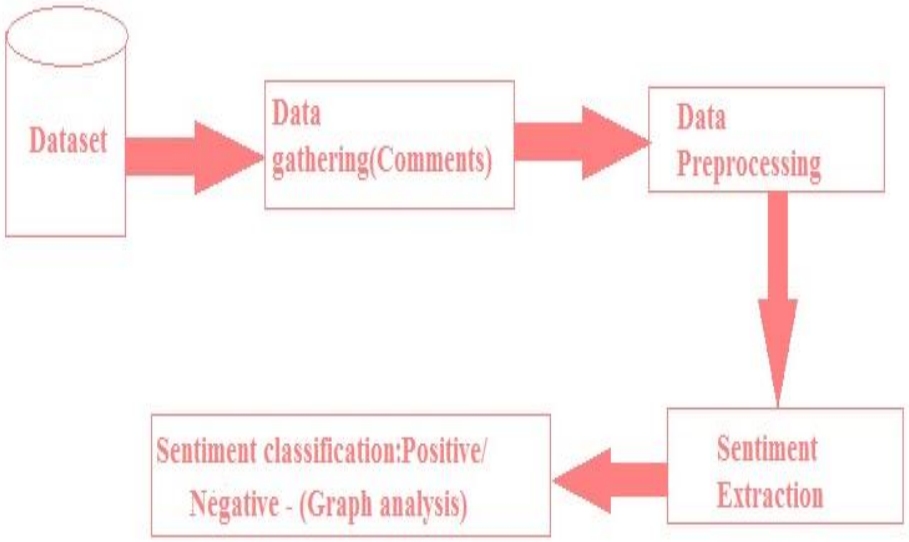

Fig 4 : Working of System

- **Preprocessing involves 3 steps:**
  - **Tokenization or segmentation**:

It is the process of splitting a string of written language into its words. Text data consists of block of characters referred to as tokens. So the documents are being separated as tokens and have been used for further processing [4].

  - **Removal of stop words**:

Stop words are the words which are needed to be filtered i.e. may be before or after natural language processing. Stop words are words which contain little informational. Various tools specifically avoid to remove these stop words in order to support phrase search. Several collections of words can be chosen as stop words for any purpose. Some search engines, removes most of the common words which include lexical words such as "want" from a text in order to improve performance. Search engine or natural language processing may contain a variety of stop words. It includes English stop words such as “and”, “the”, “a”, “it”, “you”, “may”, “that”, “I”, “an”, “of” etc. which are considered as ‘functional words’ as they don’t have meaning. Researchers have shown that by removing stop words from the file, you can get the benefit of reduced index size without much affecting the accuracy of a user's. But care should be taken however to take into consideration the user's needs. Mostly, all search engines helps in eliminating the stop words from their indexes. With the help of eliminating stop words from the index, the index size can be reduced to about 33% for a word level index. While assessing the content of natural language processing, meaning of word can be conveyed more clearly by removing the functional word [6] .

  - **Stemming:**

It is the term which used to describe the process to reduce derived words to their origin word stem. Since 1960s, algorithms for stemming have been studied in the field of computer science. Different Stemming methods are commonly referred as stemming algorithms or stemmers. For English, the stemmer example are that, it should identify the string “cats”, ”catty” as based on the root word “cat”, and also “walks”, “walked”, “walking” as based on the root word "walk" [7] .

3.3 Sentiment Extraction

It provides valuable things from text mining so that it can provide information that helps in improving decision and processes. It includes ways such as sentiment analysis, document imaging, fraud analysis etc [15]. After removing of noisy data from the only important and useful data for sentiment analysis is extracted in this section.

3.4 Sentiment & graph analysis

Data classifying and identifying is all about to tag the data so it can be create quickly and efficiently. But various organizations can gain from re- transforming their information, which helps in order to cut storage and backup costs, with increasing the speed of data searches. Classification can help an organization to meet authorized and regulatory requirements to retrieve specific information within a specific time period, and this is most important factor behind implementing various data classification technology [6].After successful implementation of sentiment analysis the result of post is shown in form of graph which shows the positive and negative nature of comments. The graph will show the respective post of user is positive or negative.

IV. FUTURE SCOPE

As current system is dealing with only textual data, in future one can collect large related data which includes videos ,documents, audios to know the sentiment analysis of user's post from multiple social networking sites which may provide better results by overcoming the limitations of the project with better improvement . User will also able to comment in any language which is limitation of current system [7]. The graph shown after the sentiment analysis will include the content of neutral nature of comments along with positive and negative nature.

IV. CONCLUSION

Social media Monitoring has been growing very rapidly so there is a need to analyze customer behavior or attitude or opinion of people on various advertisements, articles, posts, news etc and can checks the attitude of User and their opinion on particular post, product, movie etc .So, the concepts of sentiment analysis have been introduced.

We Proposed a system that determines the sentiment of the text . To perform sentiment analysis , we collected comments from users post from various social networking sites and stored it in our database as our dataset. We preprocessed the collected data by filtering out the noisy data and stop words from the data . We applied sentiment extraction to derive usefulness about stored information and removed the unnecessary contents from data .After removing unnecessary data we identified and analyzed useful data which is important for sentiment analysis by comparing it with stored keyword library, then by using a web application, the final results are shown in the form of graph on users post , whether it is positive or negative, which is extended to strength of polarity and analyzed the overall sentiment for each object by computing the weighted average for all the sentiments in the textual data.

REFERENCES

[1] Curator. S Social media: a rich source of customer sentiment for you to mine. Whatech Channel
[2] Koweika A.,Gupta A.,Sondhi K Sentiment analysis for social media. International Journal of Advanced Research in Computer Science and Software Engineering.
[3] Bissattini C., Christodoulou K. Web sentiment analysis for revealing public opinions, trends and making good financial decisions. Journal of Advanced Research in Computer Science and Software Engineering.
[4] Tulankar S.,Athale R.,Bhujbal S Sentiment analysis of equities using data mining techniques and visualizing the trends. International Journal of Computer Science Issues..
[5] Qiu M.,Yang L., Jiang J. Mining user relations from online discussions using sentiment analysis and probabilistic matrix factorization. Proceedings of NAACL-HLT, Atlanta, Georgia.
[6] Buche A., Chandak M.B., Zadgaonkar A. Opinion mining and analysis: a survey. International Journal on Natural Language Computing(IJNLC).39-48.
[7] Cataldi M., Ballatore A., Ilaria T..Good location, terrible food: Detecting feature sentiment in user-generated reviews. International Journal of Social Network Analysis and Mining (SNAM).
[8] Manjaly J.S. Twitter based sentiment analysis for subject identification. International Journal of Advanced Research in Computer and Communication Engineering.
[9] Zuell B., Preradovic N.M., Methods and usage of sentiment analysis in the context of the TV industry. International Journal of Recent Advances in Information Science..

[10] Jagtap V.S., Pawar K. Analysis of different approaches to sentence-level sentiment classification. International Journal of Scientific Engineering and Technology. 164-170. Mullen T. and Malouf R.,Taking sides: User classification for informal online political discourse. Internet Research, 2008.
[11] Nakov P.,Kozareva Z., Ritter A.,"Sentiment analysis in Twitter. Second Joint Conference on Lexical and Computational Semantics". Atlanta, Georgia.
[12] Younggue B.,Hongchul L.,"Sentiment analysis of Twitter audience: Measuring the positive or negative influence", Journal of the American Society for Information Science and Technology.
[13] Savage N. ," Twitter as medium and message", Communications of the ACM Society, Issue: 3, Vol.54. pp: 18-20.
[14] ZHU Jian, XU Chen, and WANG Han-shi, 2010. Sentiment classification using the theory of ANNs, The Journal of China Universities of Posts and Telecommunications, 17(Suppl.): pp. 58–62.
[15] X. Ding, B. Liu, and P. S. Yu,. A holistic lexicon- based approach to opinion mining, Proceedings of the Conference on Web Search and Web Data Mining (WSDM).
[16] A.Pappu Rajan , ," A Study on Security Threat Awareness among Students Using Social Networking Sites, by Applying Data Mining Techniques", International Journal Of Research In Commerce, IT & Management, Vol. No. 3 , Issue No. 09 : ISSN 2231- 5756
[17] Kwak, H., Lee, C., Park, H. &Moon, S. What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World Wide Web, 591– 600 (2010).