Sentiment Analysis for Facebook and Twitter Data

Aniket Kumar[1], Akshit Kocher[2] and Jyotika Pruthi[3]

[1,2,3]*The Northcap University*
E-mail: [1]aniketkumarkt@gmail.com, [2]kocher.akshit456@gmail.com,
[3]jyotikapruthi@ncuindia.com

Abstract—*Social networking sites such as Facebook, Twitter have become one of the important platforms for the interaction of people in day to day life. On these platforms, people express their views every day through posts, comments etc. In this paper, a survey is conducted to perform the sentiment analysis through the posts or comments and categorize them into positive (happy), negative (unhappy) and neutral (emotionless) emotion. After analyzing and categorizing all the posts, a study is done and the data is collected based on the general behavior of the people according to a post. The goal is to find out the best things that people like and what they dislike through which organizations will know how they have to design their products which can satisfy the people around the globe which will definitely result in better user experience. A company can also monitor the activity and issues people are facing to improve their services.*

1. INTRODUCTION

These guidelines include complete descriptions of the fonts, spacing, and related information for producing your proceedings manuscripts. Please follow them and if you have any questions, direct them to the production editor in charge of your proceedings at the IEEE Computer Society Press: Phone (714) 821-8380 or Fax (714) 761-1784.

We have seen an incredible development in social media over a past few years. Social media platforms like Facebook and Twitter have really revolutionized the industry and are still growing at phenomenal rate. People like to share views in forms of posts or comments on these platforms. There are also many studies that tell about the behavior of users.

This paper provides us the methods by which we can do the sentimental analysis and proposes the use of those discoveries in the business. Most of the people love to increase their likes and comments but by the use of software's we can extract the positivity or the negativity of the posts and then use the analysis to monitor the issues of related people ultimately improving the services.

This paper has been divided into 5 sections: Section 1 contains information about the related work while Section 2 discusses about the methods of extraction of data of Facebook and Twitter, Section 3Analyzes the information and section 4representation of data and section 5 concludes the research.

1.1 Related Work

For Analysis of people's behavior on these platforms, many methods have been discovered. Adamic and Adar[1] have tried to use the data from the internet about the social platforms. They broadly have analyzed the data and found that there are some aspects which can explain the social connection in a much better way. Akram Al-Kouz Et Al. [2] analyzed social graphs which are definite and accurate to know about the interests of the people and the fields where they are expert. Users are classified on the basis of groups and posts so that to prepare a more precise model of users fields of expertise. Catanese et al. [3] have developed a tool that can analyze the characteristics of a social media platform which makes it easier to understand the reasons why companies like Facebook has considering the use of techniques like BFS(Breadth- First-Search) algorithm. Ereteo et al. [4] have developed software that can use the graphical representation of social networks. Bernhard Rieder [5] has proposed a method by which one can analyze the groups and posts on Facebook. Amparo E. Cano et al. [6] developed a different Page Rank algorithm that can analyze the topics and entities by which a certain user get influenced and this data is primarily based on the retweets on Twitter. Facebook Graph APIs [7] and Facebook Query Language (FQL) Table Reference [8] are Facebook Crawlers that are used to fetch the data about the posts shared by a certain User.

Linguistic Inquiry and Word Count (LIWC) [9] is used to estimate the use of different words by the user in posts and comments. N. Azmina M.Zamani[10] work focuses on quantifying Facebook sentiments by using a lexicon based approach. Diego terrane [11] analyzed the page of a user on Facebook and discussed shares and posts with other Facebook users related to each other on various topics.

By using all the methods a sentiment analysis is done and organizations and businessman use this information to improve their services and ultimately leads to better user experience and industry growth.

2. DATA EXTRACTION

There are various methods to extract data and then analyze it. Some of them are new software's or algorithms as other uses data present on the internet to create better representations of the data in forms of graphs etc. The procedure we are using to analyze the user's profile on Facebook and Twitter is as follows:

For Twitter

Collection of tweets → Identification of language → Identification of emoticon pattern (positive or negative) → Pre-processing → Word polarity Calculation → Word polarity Dictionary→ Polarity calculation → Creation of Reports(in terms of graphs etc.).

For Facebook

Facebook crawler→ Extraction of information→ Category Identification → Polarity calculation→ Creation of Reports (in terms of graphs etc.)

In case of Twitter, the system makes the Sentiment Dictionary by diving the emoticons in positive and negative.

In case of Facebook, a crawler is used to extract the data and then category identification is done in which the system computes the extents of different words users use. To categorize them by sentiments Polarity calculation module comes to work as it classifies them into positive, negative or neutral. Creation of reports module creates graphs from the stats collected. Let's discuss the modules of procedure in detail on following subsections.

2.1 Formation of Sentiment Dictionary

The formation of the dictionary is fully automatic and focuses on the emoticons used in tweets to identify the word polarity score and then produce a sentiment lexicon of that word. Tweets in which emoticons are used has been collected through Twitter API's and then polarity is set that is it is positive or negative. If the tweets have a higher frequency of positive class words then negative, it is termed as high polarity score and similarly, if the tweet has the high frequency of negative class words then positive, it is termed as low polarity score. Let the word be m of which we want to calculate polarity so,

where,

Polarity(m)=

positive(m)-(negative(m))÷positive(m)+(negative(m))
…….(1)

Positive (m) = occurrence   of the word in positive set
Negative (m) = occurrence of the word in negative set

The score of polarity varies from -1 to +1 as +1 is the highest polarity and -1 is the lowest polarity. To collect the tweets and

queries in real time, the twitter4j [12] JAVA library has been used.

2.2 Facebook Crawler

Facebook Graph API's and Facebook Query Language (FQL) have been used to collect the data of the user from Facebook. For fetching real-time queries Facebook Graph API's written in JAVA has been used. An access token is required to access the information from Facebook as it contains set of permissions that are given to a certain organization or a person.

2.3 Extraction of Information

The above module helps in extracting the information of data from Facebook. The information is about music, songs, movies, books, posts photos, likes, comments, shares, family, activities, interests, status. Though Facebook imposes some limitation on the access of data the sources have managed to increase the accessibility from 25 likes to 100 likes by using "ad hoc".

3. ANALYZING THE INFORMATION

3.1 Category Identification

The extracted information (text) has been analyzed by the software Linguistic Enquiry and Word Count (LIWC).

This software searches for the match of the extracted word in the dictionary and then increases its category scale.

3.2 Polarity Calculation

This system uses the Sentimental Dictionary Creation module to point out the topics user has been discussing.

As we know that people don't always use dictionary words or syntax, so we have filter out the texts which contain less than k characters. To calculate polarity, a certain threshold score has been set and if the polarity score is greater than the threshold then it has positive polarity similarly if the score is less than the thresholds for negative polarity then it is negative. If the score is between both the threshold values, then it is considered to be as neutral.

To calculate the polarity of a sentence(s) we use,

where,

$$p(s)=(\sum p(n))\div k \qquad\qquad ….(2)$$

p(n)= polarity of word(n)

k= number of words in the sentence

So if,

$p(s) \geq pavg+$

Then it is considered as positive.

$p(s) < pavg+$ and $p(s) > pavg-$

Then it is considered as neutral.

p(s) <pavg-

Then it is considered as negative.

where,

pavg-mean of the polarity(all words in lexicon).

Every post on Facebook has a unique ID which one can obtain from URL bar. So after all the classification process, words get classified based on two emotion libraries which were created earlier. The counter increases on every matched word based on its type (happy or unhappy).

4. REPRESENTATION OF ANALYZED DATA

According to stats of certain user on various topics, system creates a graphical representation of data. The system creates the graph for each topic S= (N, E) where N is the number of nodes (Users) and E denotes the number of edges i.e., other user that are connected through posts to the main user. Edge represents sharing of posts between nodes (Users) and is also related to number of nodes

To understand the satisfaction of user for a post the percentages of happy, unhappy and emotionless words are compared. To measure accurately few posts get selected to be analyzed and result in more accurate sentiment analysis.

5. CONCLUSION

This paper proposes few methods of extraction of data from social networks and then performs the sentimental analysis. Many software's and tools are used for the extraction of data and classification into its emotion class. This helps to identify the overall reaction of a user to certain or if he/she is happy or unhappy or emotionless with the post. These results are very useful in many cases especially for business and for development of the industry. It can enhance the quality of service which will ultimately result in better user experience and more business.

REFERENCES

[1] L. Adamic and E. Adar, Friends and neighbors on the web, Social Networks 25(2001) 211-230.

[2] A.Alkouz, E.W.D.Luca and S.Albaryak, Latent semantic social graph model for expert discovery in facebook in IICS, 2011,pp.128- 138.

[3] S. Catanese, P. De Meo, E. Ferrara, G. Fiumara and A. Provetti, Extraction and analysis of Facebook friendship relations, in Computational Social Networks(Springer, London, 2012), pp. 291-324.

[4] G. Ereteo, MBuffa, F.Gandom and O. Corby, Analysis of a real online social network using semantic web frameworks, in International Semantic Web Conference, 2009, pp. 180-195

[5] B. Reider, Studying Facebook via data extraction: the netvizz application, in WebSci, ACM,2013,pp.346-355.Available:http://dblp.uni-trier.de/db/conf/websci/websci2013.html#Reider13

[6] A.E. Cano, S. Mazumdar, and F. Ciravegna, Social influence analysis in microblogging platforms- a topic-sensitive based approach, Semantic Web Journal 2014, pp. 290-296.

[7] and [8]Facebook graph APIs, facebook query language- resources available at https://developers.facebook.com/docs/reference/fql/.

[8] Linguistic inquiry and word count-resources available at https://www.liwc.net/.

[9] N.Amina M.Zamani ,SitiZ.Z.Abidin, Nasiroh Omar and M.Z.Z.Abidensentimental analysis :determining people's emotions in facebook (University Tekonologi Mara ,Malaysia)

[10] Diego Terrana, Agnese Augello and GiovanniPilato user relationships through sentiment classification.International journal of semantic computing Vol-8 No .3(2014)301-307(www.Worldscientific.com)

[11] twitter4j –a java library for the twitter API-resources available at http://twitter4j.org/en/index.html/