e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

Classifying the Emotions of Users Towards Software Products Using Sentiment Analysis

Tirunagiri Kavitha, Samatha Konda
Assistant Professor, Department of CSE, Sridevi Women's Engineering College, Village VattiNagulaPally, Mandal Rajendra Nagar, District RangaReddy, Telangana, India.

Assistant Professor, Department of IT, Sridevi Women's Engineering College, Village VattiNagulaPally, Mandal Rajendra Nagar, District RangaReddy, Telangana, India.

Abstract:- *Twitter empowers programming designers to track clients' responses to recently discharged frameworks. Such data, frequently expressed in the form of raw emotions, can be leveraged to enable a more educated programming discharge process. Nonetheless, naturally catching and translating multi-dimensional structures of human feelings communicated in Twitter messages is not a unimportant errand. Difficulties originate from the size of the information accessible, its naturally scanty nature, and the high level of domain specific words. Roused by these perceptions, in this paper we exhibit a preparatory report went for identifying, grouping, and translating feelings in programming clients' tweets. A dataset of 1000 tweets examined from an expansive scope of programming frameworks' Twitter sustains is utilized to lead our investigation. Our outcomes demonstrate that administered content classifiers (Naive Bayes and Support vector Machines) are more exact than broadly useful opinion investigation procedures in distinguishing general and specific feelings communicated in programming pertinent Tweets.*

We apply Bayes methods to classify the emotions from the tweets and filter them as per the categorised words of emotions. Minig is used to filter and alter the data as per required order in order to archive so we also uses regular expressions to figure out the different emotions of the dataset.

I. INTRODUCTION

Comments, reviews and opinion of the people play an important role to determine whether a given population is satisfied with the product, services. It helps in predicting the sentiment of a wide variety of people on a particular event of interest like the review of a software product their opinion on various topic roaming around the world. These data are essential for sentiment analysis [2]. In order to discover the overall sentiment of population, retrieval of data from sources like Twitter, Facebook, Blogs are essential. For the sentiment [5] analysis, we focus our attention towards the Twitter, a micro-blogging social networking website. Twitter generates huge data that cannot be handled manually to extract some useful information and therefore, the ingredients of automatic classification are required to handle those data. Tweets are unambiguous short texts messages that are up to a maximum of 140 characters. By the use of Twitter, millions of people around the world to be connected with their family, friends and colleagues through their computers or mobile phones. The Twitter interface allows the user to post short messages and that can be read by any other Twitter user. Twitter contains a variety of text posts and grows every day. We choose

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

Twitter as the source for opinion mining simply because of its popularity and data mining. The Existing Database is not able to process the big amount of data within specified amount of time. Also, this type of database is limited for processing of structured data and has a limitation when dealing with a large amount of data. So, the traditional solution cannot help an organization to manage and process unstructured data. With the use of Big Data technologies like R Programming, Python, Hadoop is the best way to solve Big Data challenges.

II. LIMITATIONS OF AVAILABLE SYSTEMS AND TOOLS FOR ANALYTICS

The limitations of available systems are not sufficient to deal with the complex structure of the big data. In this section, we present some of the limitations that are present in the existing system.

1) The available systems like Twitter-Monitor and Real Time Twitter Trend Mining System require extensive data cleaning, data scraping and integration strategies that will ultimately increase the overhead [9].

2) For real time analytics, the available system is inefficient.

3) It is very time consuming process to analyze the huge amount of data in a short period of time.

The proposed method helps to eliminate all the drawbacks mentioned above.

III. R PROGRAMMING

R is an object-oriented programming language that is a variation of the S language, and was written by Ross Ihaka and Robert Gentleman (hence the name R),the R Core Development Team,and an army of volunteers. R is a programming language and software environment for statistical computing and graphics supported by the R Foundation for Statistical Computing. The R language is widely used among statisticians and data miners for developing statistical software anddata analysis.[5] Polls, surveys of data miners, and studies of scholarly literature databases show that R's popularity has increased substantially in recent years. R is very much a vehicle for newly developing methods of interactive data analysis. It has developed rapidly, and has been extended by a large collection of packages. However, most programs written in R are essentially ephemeral, written for a single piece of data analysis.

IV. METHODOLOGY

Social networking sites acquired immense popularity and interest with the people around the world. Twitter is one of the effective tools for any business intelligence to get information about what people are talking and reacting about the topics that are roaming around the world. A twitter helps to engage the users and directly communicates with them and in response, users to provide word-of-mouth marketing for companies by discussing the product quality. With the limited resources and knowing about no one can target directly to the destination consumers, the business intelligence can be more efficient in their policy of marketing by being very selective about consumers choice they should reach out to. Fig.1. shows the steps involved in processing of twitter data.

A. Fetching Twitter Data using Twitter API:

Develop a twitter API for downloading the tweets. The Twitter API directly communicates with the Source and Sink. The Authentication keys and tokens are established that helps in communication over Twitter Server. The source is twitter account and the sink is R Buffer where all the tweets are stored.

B. Pre-processing of tweets:

The data coming out from twitter contains various nonsentiment contents such as website link, emoticons, white spaces, hashtag etc. which should be removed before processing it so that the sentiment generated are accurate. Preprocessing includes:

1) Removal of URL's: Twitter data consists of different type of information. If any user posted any link which is none of the use for sentiment analysis. Therefore, URL should be removed from the tweet.

2) Removal of special symbol: There are various types of

symbols used by the user such as punctuation mark (!), full

stop (.) etc. which does not contain sentiment. Therefore,

special symbols should be removed from the tweet.

3) Converting emotions: Table. I. shows the various emotions used for conversion. Nowadays emotions become a way for the user to express their views, feeling, and emotion.Emotions play a big role in the sentiment analysis. Therefore, convert the whole emotions into its equivalent word by which we can do the analysis efficiently.

4) Removal of Username: Every Twitter user has a unique username, therefore, anything is written by a user can be indicated be writing their username proceeding by @. This type is denoted as proper nouns. For example, @username.

This also has to be removed for effective analysis.

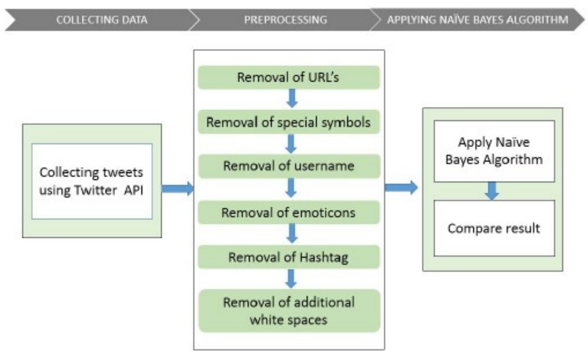

Fig 1: Architecture diagram

5) Removal of Hashtag: A hashtag is a prefixed with the hash symbol (#). Hashtag are used for naming subjects or phrases that are currently in trend. For example, #google, #twitter.

6) Removal of additional white spaces: There may be consists of extra white space in the data and it needs to be removed. By removing white spaces the analysis to be done more efficiently.

C. Applying Naive Bayes Algorithm:

The Naive Bayesian Classification [7] represents a supervised learning method as well as a statistical method for classification. It is probabilistic model and it permit us to capture uncertainty about the model in a principled way by determining probabilities. It helps to solve diagnostic and predictive problems. This Classification is named as Naive Bayes after Thomas Bayes, who proposed the Bayes Theorem of determining probability. Bayesian classification provides useful learning algorithms and past knowledge and observed data can be combined. It helps to provide a useful perspective for understanding and also evaluating many learning algorithms. This helps to determine exact probabilities for hypothesis and also it is robust to noise in input data.

International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)}$$

P(C | X) is posterior probability,

P(X | C) is likelihood,

P(C) is class prior probability,

P(X) is predictor prior probability.

V. FRAMEWORK IMPLEMENTATION OF NAÏVE BAYES ALGORITHM

In this section, we present the implementation of our Hadoop framework for efficiently executing Naive Bayes algorithm. Our proposed mechanism extends Hadoop to implement map and reduce phase. To implement Naïve Bayes algorithm we need a trained SentiWordNet [1] dictionary which is available online. It consists of collection of different word with its synonym and its polarity. The synonym represents the similar word meaning which will be having same polarity. The polarity represents the positivity of the word in the context of the sentence.

Sentiments	Count without emotions	Count with emoticons
Extreme Positive	130	177
Positive	59	90
Extreme Negative	45	42
Negative	30	26
Neutral	136	65

Table: Polarity of the data

VI. RESULT

Firstly the data is downloaded from the twitter. They are stored into the R buffer for analysis. Before evaluation the tweets, we need to be pre-processed in order to remove the noise from the data. We evaluate performance of our algorithm by comparing the result with and without considering emotions. We observed in our result that when we perform pre-processing without considering emotions the tweets. contains sentiment in the form of emotions are simply ignored by Naive Bayes algorithm and hence when we perform preprocessing by considering emotions, the results are much more accurate than the previous one. From the above table, we observed that the sentiments which is neutral are in great number while emotions are not

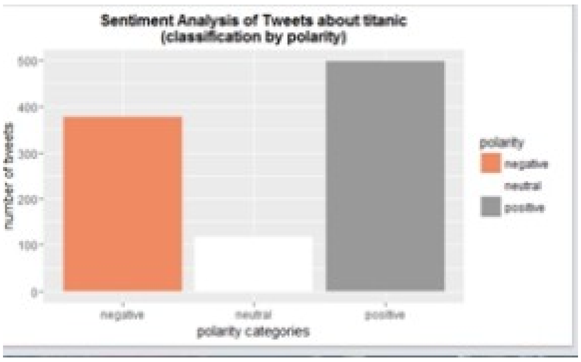

Fig 2: Polarity of the data set

Considered. But when pre-processing is used with emotions the neutral tweets are considerably decreases due to conversion of emotions. Hence performance of Naive Bayes algorithm increases by converting the emotions by assigning its equivalent word.

VII. CONCLUSION AND FUTURE WORK

Twitter Data in the form of opinion, feedback, reviews, remarks and complaint are treated as big data and it cannot be used directly. These data first convert as per requirement. In this paper, we discussed pre-processing of data to remove noise from the data. We have implemented sentiment analysis for Software Product set, on R programming and analyzed with large number of tweets. This type analysis will definitely help any organization to improve their business productivity. The analysis of twitter data are done on various perspective like Positive, Negative and Neutral sentiments on tweets. It also provide the fast downloading approach

e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue14
November 2017

for efficient Twitter Trend Analysis. Tweets can also be useful in prediction of product sales, quality of services offered by company, feedback of users etc. Hence, the future scope in the sentiment analysis for the other social networking websites like Facebook, Google Plus etc.

VII. REFERENCE

[1] F. Zhu and X. Zhang, “Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,” Journal of marketing, vol. 74, no. 2, pp. 133–148, 2010.

[2] Q. Ye, R. Law, and B. Gu, “The impact of online user reviews on hotel room sales,” International Journal of Hospitality Management, vol. 28, no. 1, pp. 180–182, 2009.

[3] Q. Roller, “Quote roller reviews, pricing comparison of alternatives,” GetApp, vol. 02-, 2015. [Online]. Available: http://www.getapp.com/sales-oftware/a/quote-roller/

[4] N. Mishra and C. Jha, “Classification of opinion mining techniques,” International Journal of Computer Applications, vol. 56, no. 13, pp. 1–6, 2012.

[5] S. Model, “Software as a service (saas) from the gartner IT glossary,” Gartner IT Glossary, vol. 04-, 2012. [Online]. Available: http://www.gartner.com/it-glossary/software-as-a-service-saas/

[6] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” Foundations and trends in information retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.

[7] D. Song, H. Lin, and Z. Yang, “Opinion mining in e-learning system,” in Network and Parallel Computing Workshops, 2007. NPC Workshops. IFIP International Conference on. IEEE, 2007, pp. 788–792.

[8] E. Marrese-Taylor, J. D. Vel´asquez, F. Bravo-Marquez, and Y. Matsuo, “Identifying customer preferences about tourism products using an aspect-based opinion mining approach,” Procedia Computer Science, vol. 22, pp. 182–191, 2013.

[9] S. F. Sayeedunnissa, A. R. Hussain, and M. A. Hameed, “Supervised opinion mining of social network data using a bag-of-words approach on the cloud,” in Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012). Springer, 2013, pp. 299–309.

[10] N. Bharathi, G. Brindha, B. Santhi, and P. Neelamegam, “Integration of fpga, cloud and opinion mining,” Journal of Applied Sciences, vol. 14, no. 14, p. 1653, 2014.

[11] M. Cooke and N. Buckley, “Web 2.0, social networks and the future of market research,” International Journal of Market Research, vol. 50, no. 2, p. 267, 2008.