International Journal of Pure and Applied Mathematics
Volume 116 No. 5 2017, 69-74
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

IMPLEMENTATION OF SENTIMENT ANALYSIS ON TWITTER DATA

Thirupathi Rao Komati[1], Sai Balakrishna Allamsetty[2], Chaitanya Varma Pinnamaraju[3]
[1]Professor, Dept. of CSE, K L University, India.
[1]kthiruapthirao@kluniversity.in.
[2]Student, Dept. of CSE, K L University, India.
[2]allamsettysaibalakrishna@gmail.com.
[3]varmachaitanya420@gmail.com.

Abstract: As the data being generated is growing rapidly at a scale petabytes per day in various forms one of the main sources that generates such large data is social media platforms with over tens of millions of people active per day, if only we could use this data to extract useful information for analyzing the current business needs, their reach and customer satisfaction towards the product and the company. It could contribute to meet constantly changing requirements and also analyze other competitors' performance and change one's business strategies accordingly to be on the top. In this project, we are going to take data generated by users of one of the top microblogging websites Twitter, which has over 100 million daily active users and we are going to implement sentiment analysis on the tweets. This paper produces the output in the form of graphical representation of various tweets describing the total sentiment score of the tweets and as well as it also produces the individual score of each tweet.

Keywords: Sentiment Analysis**,** Twitter

1. Introduction

R is a popular programming language which is generally embraced by information researchers. In any case, normal R must be executed in a solitary machine environment. As the volume of accessible information proceeds to quickly develop from an assortment of sources, versatile and execution investigation arrangements have turned into a fundamental device to upgrade business profitability and income. Existing information examination situations, for example, R, are compelled by the span of the fundamental memory and can't scale in numerous applications. Information representation is turning into an undeniably imperative part of investigative in the time of huge information .

[1]The steps required for analyzing data are:

- The need for Meeting speed
- Understanding the various types of data
- Addressing the data quality
- Display the correct understandable results

The R environment provides enormous built-in functions in the package "base", most of which are generally required for elementary data analysis (e.g., linear modeling, graph plotting, basic statistics). However, the real beauty of R is its almost versatility and infinite expandability. Approximately 2,500 packages have been developed for R by the active R community. [2]These bundles serve to enlarge R's regular abilities in information examination and frequently concentrate on improvements in different scientific fields, and additionally systems utilized as a part of very specific information investigations.

Sentiment analysis: It is likewise alluded to as the Opinion Mining, Which suggests separating different feelings, feelings and estimations in content. As you can envision, a standout amongst the most well-known uses of opinion examination is to track mentalities and emotions on the web, particularly to tack items, administrations, marks or even individuals.[4]An principle thought is figure out if are emphatically or adversely by a given gathering of people. The reason for Text Mining is to prepare literary data, extricate significant numeric records from the content, and, in this way, make the data contained in the content open to the different information mining calculations.[1] Data can be separated to determine synopses for the words contained in the archives or to process rundowns for

the reports in view of the words contained in them. Henceforth, you can investigate words, bunches of words utilized as a part of archives, or you could break down records and decide likenesses between them or how they are identified with different factors of enthusiasm for the information mining venture. In the most broad terms, content mining turns "content into number" which can then be joined in different examinations. Utilizations of content Mining are breaking down open-finished study reactions, programmed preparing of messages, messages, and so on., dissecting guarantee or protection claims, demonstrative meetings, and so forth., researching contenders by slithering their sites.

The packages used in this project are:

- twitteR **:**It is an interface to access Twitter API.
- plyr**:** This package is a collection of tools which can solve general set of problems like when we need to break down a large problem to various pieces and each piece is operated separately and then all the pieces put back together.
- Stringr**:** It is a collection of simple wrappers which help us in doing basic operations on strings like removing special characteristics , converting uppercase alphabets to lower case alphabet and by not considering spaces.
- ggplot2 **:** Used to plot graphics using grammar in R. Each plot can be build up step by step from various different sources

2. Literature survey

The paper "Sentiment analysis of twitter published in 2012 introduces a machine learning approach to implement sentiment analysis on the data. [5]They have performed sentiment classification of Twitter data where the classes are positive ,negative and neutral. Two sorts of models have been utilized: Tree part and highlight based models and both these models beat the unigram pattern. For the element based approach they performed include examination, Which uncovers that the most critical components are those that join the earlier extremity of words and their parts of discourse labels.

In "The Twitter Sentiment Analysis: The Bad The Good and The OMG" paper, they have explored the utility of phonetic components for recognizing the assumption of twitter messages.[8] They have known the value of existing of lexical assets and in addition includes that catch data about the casual and imaginative dialect utilized as a part of various social sites. An approach has been introduced to solve the problems.

In This Paper "The Twitter Sentiment Classification using Discrete Supervision" published in 2009 introduces a novel approach for naturally grouping the feeling of various twitter message.[7] These messages are either classified as The positive or negative with respect to the data . The paper describes the preprocessing of various steps in order to achieve very high accuracy. The principle commitment of this paper is to utilizing tweets with emoticons for far off regulated learningDiverse machine learning classifiers and highlight extractors have been utilized alongside the utilization of unigrams, bigrams, unigrams and bigrams, and parts of discourse as components

3. Proposed Work:

We have proposed a system that performs aspect level sentiment analysis on twitter data or tweets based on movies into two categories:

- Positive
- Negative

[6]The Following is a brief Description associated with the various tweets.

- Emoticons: The expressions which are used to describe the users conditions or feelings for an issue or his personal issues.
- Target: The Twitter users will use the special characters "@" symbol to simply refer to other users on the various micro blog which continuously alerts them
- Hashtags: The Users usually use The hash tags (#) to refer to various topics. This is to increase the views of their personal tweets.

Aspect Level Sentiment Classification:

Sentence level or document level sentiment classification is insufficient in many applications as it only reflects the overall opinion and does not evaluate all the aspects of an entity. Hence in order to understand the sentiment of each aspect, [11]We perform aspect-level sentiment analysis or feature-based opinion mining. This paper, proposes to

perform sentiment analysis of multiple aspects of various entities related to movies, products, companies.

For example:

- "I love #hrithik so much, cant wait to see his film"

When we want to find the tweets above the hero Hrithik. Let us consider the above tweet as the retrieved data . Now we apply the sentiment function to the above tweet.The word "Love" in the above tweet is a positive word . So the score of the tweet would be +1.

- "I abhor @hrithik movies.

When we want to find the tweets above the hero Hrithik. Let us consider the above tweet as the retrieved data . Now we apply the sentiment function to the above tweet.The word "abhor" in the above tweet means negative word. So the score of tweet would be -1.

- " I love #hrithik so much, but I abhor his movies."

Let us consider this tweet as the retrieved data , Now let us apply the sentiment function on the above tweet.The word "Love" and "abhor" are positive and negative words in the above tweet. So the score of the tweet would Zero.

The following steps proposed in our paper are :

- Data collection using Twitter API: Large sets of twitter data is not available publicly. Hence we first extract the twitter data from the Twitter API.
- Data Preprocessing: This involves cleaning and simplifying the data by performing spell correction, punctuation handling etc. so as to remove noise from the data.
- Applying Classification algorithms: The classification algorithms are applied on these tweets in order to categorize them. Different models provide different accuracy and we choose the model with highest accuracy.
- Classified tweets: The results of the above step is classifies tweets which may belong to any of the three categories mentioned.
- Sentiments in graphical representation : The results of the sentiment analysis is provided using histograms

4. Implementation

The steps in implementation are :

a) Load Twitter API
b) Load word dictionaries
c) Search twitter feeds
d) Getting text from feeds
e) Defining text cleaning functions
f) cleaning and splitting twitter feeds
g) Analyzing twitter feeds
h) Plotting high frequency negative and positive words

a)Load Twitter API :

The first step is to register in the twitter application developer's portal and get the authorization. You need :

consumer_key<- "Your Twitter Consumer key"

consumer_secret<- "Your Twitter Consumer Secret key"

access_token<- "Your Twitter Access Token key"

access_secret<- "Your Twitter Access Secret key"

b) Load word dictionaries :

Next stride is to stack the arrangement of positive and negative assumptions words into your R working catalog. The words are then gotten to and relegated to factors, positive and negative as demonstrated as follows.

c) Search twitter feeds:

The following stride is characterizing a twitter seek string and relegating to a variable, Number of tweets to be removed is alloted to another variable, number. An ideal opportunity to play out the twitter hunt and extraction is influenced by this number. A moderate web association as well as unpredictable inquiry fields may bring about extra postponements.

d) Getting text from feeds :

Twitter sustains have huge amounts of extra fields and implanted pointless data. We utilize the gettext() capacity to remove the content fields and appoint the rundown to a variable tweetT. The capacity is connected to every one of the 5000 tweets. The code beneath likewise indicates

consequences of extraction for the initial 5 sustains.tweetT=lapply(tweet,function(t)t$getText())

head(tweetT,5)

e) Defining text cleaning functions :

In this progression, we compose a capacity which executes a progression of orders to clean content, removes punctuation, special characters, inserted HTTP joins, additional spaces, and digits. This function likewise changes capitalized characters to lower case string utilizing tolower() work. Ordinarily, the tolower() work stops startlingly as it experiences unique characters ceasing execution of the r code. To dodge this, we compose a blunder getting capacity, tryTolower, and install it in the code of the content cleaning function.

f) Cleaning and Splitting twitter feeds :

In this step, we separate the tweets. The resultant feeds are stored in a list object.

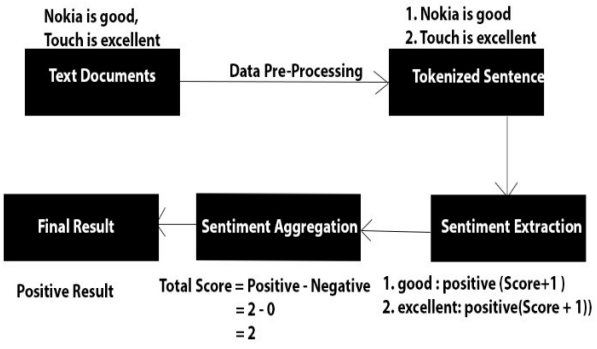

Figure 1. Sentiment analysis

g) Analyzing twitter feeds :

Here we get into the actual task of analyzing feeds. We compare the twitter text feeds with the word dictionaries and retrieve out the matching words. To do this, we first define a function to count the positive and negative words that are matching with our database.

h)Plotting high frequency negative and positive words:

The resultant output for the word hrithik is

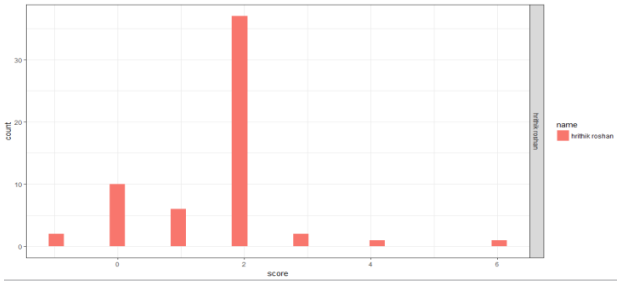

Figure 2. Graphical representation

The tweets of Figure2 graph are:

score	text	name	code
1	@hrithik when is the party? or is it that you only cele...	hrithik roshan	hr
0	RT @adcjasmins: do a film together i've been starvin...	hrithik roshan	hr
0	RT @adcjasmins: do a film together i've been starvin...	hrithik roshan	hr
-1	@TheLoyalSRKFan @iHrithik ..Blind fan !! Spectacular ...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
1	@BeingSalmanKhan @SrBachchan @hrithik Coming t...	hrithik roshan	hr
0	RT @Tauseef8143: #NTR27Launch TRENDING ROIIII ...	hrithik roshan	hr
0	#NTR27Launch TRENDING ROIIII @Amaresh143143 @...	hrithik roshan	hr
0	@Jeny472286651 @hrithik nazia u r mst beautiful.....	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
1	Retweeted KAABIL ROCK <ed><U+00A0><U+00BD><...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
1	@nidhira38682641 @AnkzBhargava @iHrithik @Hrithi...	hrithik roshan	hr
6	@AnkzBhargava @iHrithik @Hrithik lovely love kaabil ...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr
2	RT @_ajaytiwari79: @Hrithik sir Is Like Sun Noboby C...	hrithik roshan	hr

Figure 3. Tweets

5. Conclusion

The project helps us to analyze huge amount of data and process it. The data will be collected by the twitter streaming API. The data collected will be analyzed, based on score we analyze how users are feeling about the product or company etc. We can also use this to visualize the users' opinion towards other products in the market by drawing a bar graph. This project not only analyzes the sentiments of the users but can be very helpful in marketing sector.

References

[1] Luciano Barbosa and Junlan Feng. 2010. The sentiment detection on twitter based on the noisy data.

[2Robust sentiment detection on twitter from biased and noisy data. Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pages 36–44.

[3] The Michael Gamon. 2004. The classification of sentiment on the feedback of data retrieved: noisy data, large feature vectors, and the role of linguistic analysis.

[4] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter classification of sentiment using far supervision. Technical report, Stanford.

[5] The David Haussler. The kernels on different structures. University of California in Santa Cruz.

[6] M Hu and B Liu. 2004. The Mining of data and summarizing the customer reviews based on an issue. KDD

[7] S M Kim and E Hovy. 2004. Determining the sentimental analysis of various opinions. Coling.

[8]Erumit, A.K.Nabiyev, V. ; Cebi, A.Karadeniz Tech. University, Trabzon, Turkey, “The Design of motion problems based on The graphical theory in counting the number of words in maths .

[9]Peter D. Zegzhda, Dmitry P. Zegzhda, Alexey V. Nikolskiy, "the Graph Theory in Cloud computing System Which provides Security in Modeling", The Computer Network Security, ISBN 978-3-642-33703-1, Volume 7531, pp 309-318, 2012.

[10] Brindha.C , Murari Devakannan Kamalesh, “Smart Alert For Eb Metering With Enhanced Security”, International Innovative Research Journal of Engineering and Technology, Volume 2, pp. 70-75, 2016.

[11] Yexia Cheng; Yuejin Du; JunFengXu; Chunyang Yuan; ZhiXue, "The security evaluation on cloud computing based on the graphs which are represented in R" Cloud Computing and The 2012 IEEE second International Conference on , vol.01, no., pp.459,465, Oct. 30 2012-Nov. 1 2012.

[12] Council I, Mc Donald R, Velikovich L, “The learning of scope of negation to classify the sentimental analysis and representing in the form of graphical”