Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary | Peer Reviewed | Open Access
Volume: SI, Number: 01, Pages: 95- 99, Year: 2020

Sentiment Base Emotions Classification of Celebrity Tweets by Using R Language

Sumera Mehboob[1], Syed Ali Jafar Zaidi[1, *], Muhammad Rizwan[1], Usman Dilshad[3], Nadeem Lahsari[4], Adeel Hassan[4], Ghulam Hassan Sanwal[5]

[1]Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan- 61010, Pakistan.
[3]Government College of Technology Bahawalpur, Pakistan
[4]Islamia University Bahawalpur Baghdad Ul Jadeed Campus, Bahawalpur, Pakistan
[5]Department of Electrical Engineering, Khawaja Fareed University of Information Technology RYK
*Corresponding Author's email: ali.zaidi@kfueit.edu.pk

Abstract: **Twitter is considered as one of the most effective microblogging site(s) developed mainly to express the views and thoughts of its users. Twitter users follow their favourite personalities, i.e., celebrities, and use to tweet/retweet, frequently, without knowing the emotions behind the made tweet. This research focuses on learning the emotions behind the tweets using R language by formulating it as an emotion classification problem. We applied twitter scraper data scraping technique to collect the dataset from the twitter accounts for our analysis. By using the proposed scheme, we estimate the emotions (whether the person was happy, sad, anxious, joyous, angry, surprised or feared) behind the tweet. We believe that our scheme would help users understand more the personality insights from the tweets. SentimentR Package of R programming language is used to find out the personality insight than the positive, negative and neutral emoticon combined to find out the accurate results. If the user has more negative tweets we can say that he is happy and joys or if the negative tweets are more than positive tweets we can quickly evaluate that the personality is sad, angry, anxious or feared. The main contribution of this paper is to identify the emotions and the trend of the characters or twitter users.**

Index terms: Twitter, Personality Insights, Sentiment Analysis, Scraping, Feature Extractions

I. INTRODUCTION

With the fast growth of internet usage, people are using social network websites to express their thoughts, views, and ideas. With the help of these microblogging sites, people get information according to their interests, besides generating a lot of data in the form of texts, images, and videos. Among all microblogging forums, Twitter is considered as the most effective forum to share the thoughts in the form of tweets, immediately and effectively. The growth of Twitter users is exponential, i.e., billions of users are using Twitter across the world (see Fig. 1), to share their thoughts, emotions, and work-related content. As a result, billions of tweets are made across the world every day [1-3].

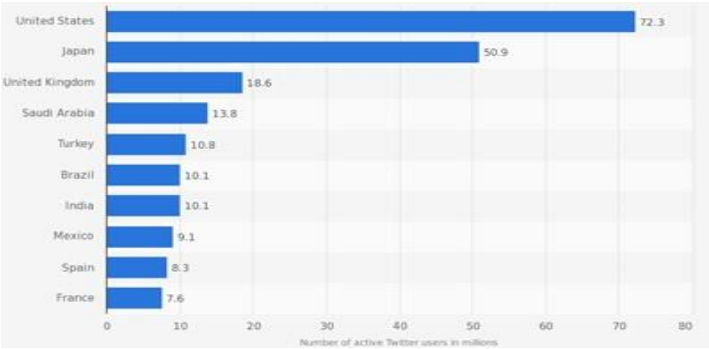


FIGURE 1: LeadingCountriesbasedonnumberofTwitter (in millions)

Millions of Twitter users follow their favourite personalities, i.e., celebrities, and tweet addressing them, or they retweet their tweets without knowing the background of their earlier made tweet. This blind retweeting sometimes has sentimental consequences [4].

This research focuses on solving the problem of the users' emotion recognition from the tweets they make. The solution also includes some challenges related to the creation of dataset and applying the machine learning classifiers [5-13] effectively. We used twitter scraper - a scraper technique, to parse the dataset as per our needs and apply sentiment analysis to estimate the users' emotions. We performed classifications in R language - as it is considered a useful language for text mining, in recent days [8]. The emotions are classified into eight different classes, namely, sad, happy, anxious, surprise, angry, joy and disgust. Tweets of different personalities of different regions were analyzed, which shows the sentiment analysis of the tweets. For this purpose, we will divide the results into three categories which are positive negative and neutral. In Positive category, we'll combine Hope and Joy tweets because it shows the positive emotions in the Negative Category we'll combine the tweets, which show disgust.

Sadness, Anger and Fear factors and in the last the only Attitude will be included. Due to the increase in the popularity of Twitter, people are using twitter on a priority basis. They want to know the emotions of those personalities which are followed by them. So we have worked on an approach to recognize the emotions of the personalities using R programming language with the help of SentimentR Module.

II. MATERIALS AND METHODS

There is a lot of research has been done on how sentiments are expressed. Automatic part of speech (POS) tags and resources such as sentiment lexicons have proved useful for sentiment analysis [1]. Researchers have also worked on detecting sentiment in the text that is called semantic orientation; hence these microblogging websites are rich sources of data for sentiment analysis and opinion mining [2, 14].

With the increase in the use of microblogging, it becomes a difficult task to extract the insights of the users. Hence, the authors applied CRF and SVM learners to identify the classifications of sentiments at the sentence level that can be related to tweets as well some researchers used the approach of Cosine Similarity and Sequence Matcher Techniques to find out the related and desired results for the sentence matching and find out the products using that technique [3, 4]. A personality refers to biophysical characteristics that uniquely define a person; few researchers have looked particularly at personality for their predictive model [8]. Sentiment Analysis and Opinion minings are one of the emerging field and computational study of people opinion. The different algorithm has also been used for the sentiment analysis and the best result for the opinion mining in the computational research of any sort of text [9, 15].

The fast development in the digital data is a problem for sentiment analysis. Today's information is composed of structured and unstructured data; text mining is considering the technique to fetch the desired results from both type of data [11-17]. To develop a model, this classifies the tweets into the complimentary negative and neutral classes. Three types of models were used for classification (i) Unigram Model (ii) A feature-based model (iii) Tree kernel-based model. The conclusion shows that uni-gram model as hard baseline achieving over 20% over chance baseline for both classifications and then they also did this with the combinations of two models like unigram model with feature and feature with tree models and they concluded that both these combinations perform the unigrams baseline for over 4% for both classification tasks [7].

A. PROBLEM STATEMENT:

This study focuses on the problem to extract the accurate, practical and related in-sights of the celebrity tweets to facilitate the other users to know about the celebrity tweets and thoughts of the celebrity and also the overall behaviour of the celerity in the tweets.

B. APPROACH

This study addresses the problem of extracting useful and effective features resulting in understanding the mood of the target celebrity. The study could be useful for the followers to understand more the contents of the tweets. Our approach starts with the tweets extraction from Twitter, using the Twitter scraper, which returns a Comma Separated Value (.csv) file containing all the tweets based on the different parameters for every single user. The retrieved tweets were not according to our requirement, so we further process the collected tweets. So that we'll be able to find the processed form of tweets and it will be easy to apply different types of experiments at this type of processed form of data.

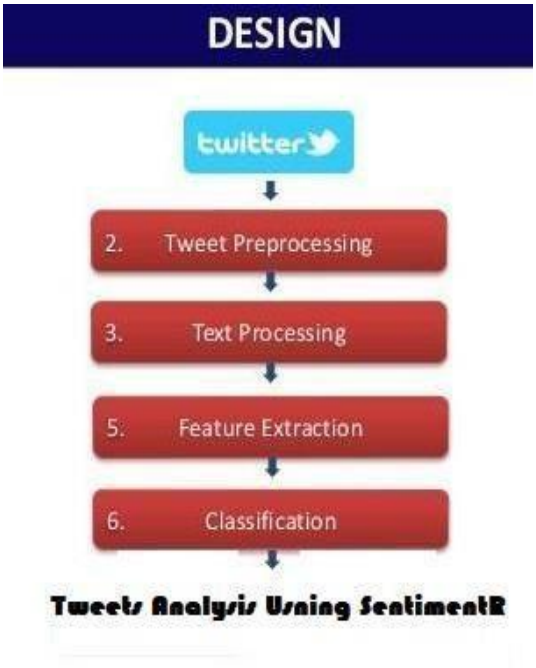


FIGURE 2: Twitter Analysis Using SentimentR

C. EMOTICONS

Finding personality insights is the main and very difficult task which depends upon the emotions. Emotions are categorized in different classes by using emotion classification algorithms. Emotions are classified in 8 classes like happy, sad, anxious, surprise, angry, fear, trust and anticipation. This classification made differentiate the emotions of a single celebrity. It tells that from which class any celebrity belongs mostly and which emotion

relates to that personality more [2]. Emoticons are based on a string of symbols that is used for the representing body language in text-based communication. In Natural Language Processing (NLP), emoticons have been considered unnatural language entities [13].

III. IMPLEMENTATION:

The systems which we have developed evaluate the sentiment analysis and give the emotion classification of tweets. For this purpose, we have used some techniques of sentiment analysis and emotion classification. This includes some steps:

i) Data collection

ii) Data filtering

iii) Sentiment Analysis

All of the above steps have been briefly described below.

A. DATA COLLECTION:

We have collected data through twitter scraper, which gives the data sets of Twitter accounts of the selected celebrities twitter accounts in the CSV file. Twitter Scraper is a module we used to scrape profiles, timelines, likes, search and tweets. It also used to extract the tweets from Twitter in your application.

id	created_at	text	retweet_count	favorite_count
771688054344089000	02-09-16 12:35	I visited our Trump Tower campaign head	2608	10070
771687247951298000	02-09-16 12:32	People will be very surprised by our grou	3152	9715
771686352438042000	02-09-16 12:28	Just heard that crazy and very dumb @m	2520	8233
771481226578460000	01-09-16 22:53	I will be interviewed by @ericbolling tor	2977	12259
771393537900503000	01-09-16 17:05	I am promising you a new legacy for Ame	8047	24078
771352519457005000	01-09-16 14:22	Thank you for having me this morning @	5441	17477
771298236376178000	01-09-16 10:46	Poll numbers way up - making big progre	8875	33909
771296597963661000	01-09-16 10:40	Thank you to @foxandfriends for the gre	6357	25783
771294347501461000	01-09-16 10:31	Mexico will pay for the wall!	26345	50213

FIGURE 3: Tweets Fetched by the twitter scraper in CSV file

How data is extracted and displayed in the CSV file is shown in Fig. 3. All the information has been fetched using twitter scrapper.

B. DATA FILTERING:

The data that was extracted from CSV file contained a lot of additional information which was not significant. So we extracted only useful data which we needed that contain a number of likes, tweets that were replied, retweets and text. Some preprocessing is also done so that we have removed stop words from it.

IV. SENTIMENT ANALYSIS

Sentiment analysis is one of the fastest and the growing research area not only in computer science but also in the field of sciences. In recent years sentiment analysis has been shifted from analyzing the online product reviews to social media Text from Facebook, Instagram and Twitter. Several other fields, like the stock market, elections, disasters, medicines, software engineering and cyberbullying, extend to the utilization of sentiment analysis [10]. We use Rstudio to developed an application to perform the sentiment analysis. We used "SentimentR" library to fetch the sentiments from the tweets. "SentimentR" package for R is beneficial when it comes to analyzing text in psychological or sociological studies. This library leverages equation 1 to perform the sentiment analysis. The equation first converts a paragraph into sentences and then turns these sentences into words. Then the stopwords are removed except paused punctuation, which is also considered as words [5]. The SentimentR package makes sentiment analysis easier as it just requires a few lines of code. This package also corrects inversions, e.g., while a more basic sentiment analysis algorithm would judge "I am not good" as positive due to the presence of an adjective (good). SentimentR recognizes this inversion and classified it as negative[6]. The equation is written below:

$$\delta_{i**j} = c'_{i**j}/\sqrt{w_{ij}n}$$

With this equation, sentiment analysis, and tokenization are also performed. The equation used by the algorithm to assign a value to the polarity of each sentence first utilizes a sentiment dictionary to tag polarized words [5]. Sentiment analysis data is tokenized and emotion classification to separate the emotions into different classes.

V. SENTIMENTR LIBRARY

The use of social media has become an integral part of our daily life. We found multiple techniques and algorithm for sentiment analysis SentimentR.

SentimentR is designed to found the text polarity at the sentence level quickly. The SentimentR package for R is immensely helpful when it comes to analyzing text for psychological or sociological studies [14]. Its first significant advantage is that it makes sentiment

analysis achievable and straightforward within a few lines of code. It is a second big advantage is that it corrects for inversions, meaning that while a more basic sentiment analysis would judge "I am not good" as positive due to the adjective good, SentimentR recognizes the inversion of interest and classifies it as negative [12].

VI. RESULTS AND DISCUSSION

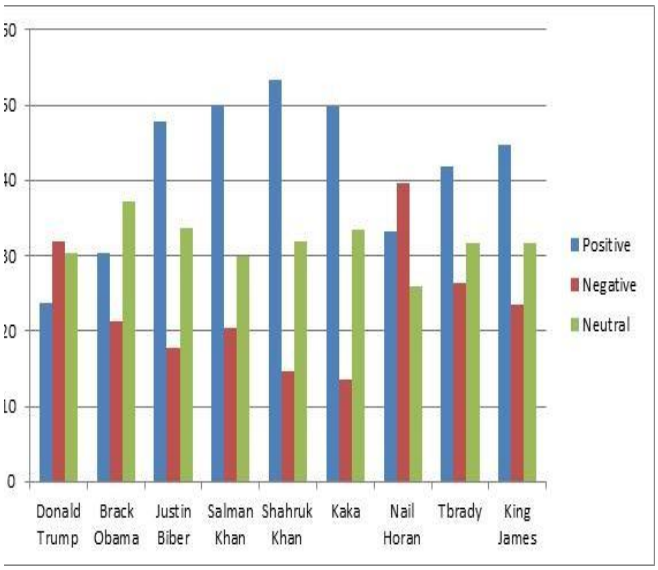
All the tweets have been analyzed using sentiment library, and then the emotions which are classified into different classes are separated now.

Twitter ID	Trust	Anger	Hope	Fear	Joy	Sad	Disgust	Surprise
Donald Trump	22.4	13.4	13.3	12.93	10.5	9.9	9.5	7.84
Barack Obama	30.7	9.9	17.5	14.08	12.8	5.6	2.7	6.60
Justin Bieber	23.4	4.8	25.8	6.29	22.0	5.3	1.9	10.29
Salman Khan	19.8	4.7	28.2	9.26	21.8	4.1	1.7	10.19
Shahrukh Khan	22.8	3.9	21.0	4.76	32.4	4.7	1.2	9.02
Kaka	26.6	4.14	25.6	7.67	24.3	3.63	1.21	6.76
Nail Horan	17.3	9.37	18.05	10.46	15.9	11.47	8.36	8.64
Tbrady	19.4	7.67	22.64	7.83	19.2	7.40	3.54	12.23
King James	23.3	6.71	23.22	7.23	21.4	6.19	3.46	8.39

TABLE 1: Sentiment analysis for different moods of the personality

In the Tweets of different personalities of different regions were analyzed, which shows the sentiment analysis of the tweets. By using this percentage, we will check the behaviour of the personality. For this purpose, we will divide the results into three categories which are positive negative and neutral. In Positive category, we'll combine Trust, Hope and Joy tweets in the Negative Category we'll combine the tweets which show Disgust, Sadness, Anger and Fear factors and in the last the only the tweets having Trust and Surprise attitude will be included.

If the sum of the percentage of Hope and joy is more than the Negative category tweets, then we'll be able to say that the person has a positive approach towards.



Graph 1 Comparison of Positive, Negative and Neutral Tweets

In Graph 1, we can easily find the positive and negative insights of the personality. As discussed earlier, the results that are presented in Figure 4 provide some new insight into Negative, Positive and Neutral aspects of the personality at a minimum. These results serve as a foundation to evaluate the twitter tweets insights of the personality. For instance, if we see the positive tweets of Salman Khan it is almost about 50% so we can say that he has a positive approach towards any model. Same in the case of Donald Trump he has more negative tweets as compared to neutral and positive tweets. In this way, we can justify the tweets and the insights of those personalities who are using twitter.

VII. CONCLUSION

In this research, we presented the case study on the insights of the personality using twitter tweets of different personalities who belong from various fields and other origins. To take advantage of a large number of statically significant words and tweets that emerges from such a large dataset. We present result from of Positive, Negative and neutral behaviour of the personality which is shown in the tweet.

REFRENCES

[1] Kouloumpis, Efthymios, Theresa Wilson, and Johanna Moore. "Twitter sentiment analysis: The good the bad and the omg!." In Fifth International AAAI conference on weblogs and social media. 2011, pp.538-541.

[2] Go A, Bhayani R, Huang L. Twitter sentiment classification using distant supervision. CS224N project report, Stanford. 2009 Dec;1(12):2009.

[3] Pak, Alexander, and Patrick Paroubek. "Twitter as a corpus for sentiment analysis and opinion mining." In LREc, vol. 10, no. 2010, pp. 1320-1326. 2010.

[4] Zaidi, Syed Ali Jafar, Attaullah Buriro, Mohammad Riaz, Athar Mahboob, and Mohammad Noman Riaz. "Implementation and comparison of text-based image retrieval schemes." International Journal of Advanced Computer Science and Applications 10, no. 1 (2019): 611-618.

[5] Hamdan, Hussam, Patrice Bellot, and Frederic Bechet. "lsislif: Feature extraction and label weighting for sentiment analysis in twitter." In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015), pp. 568-573. 2015.

[6] Agarwal, Apoorv, Boyi Xie, Ilia Vovsha, Owen Rambow, and Rebecca J. Passonneau. "Sentiment analysis of twitter data." In Proceedings of the workshop on language in social media (LSM 2011), pp. 30-38. 2011.

[7] Schwartz, Hansen Andrew, Johannes C. Eichstaedt, Lukasz Dziurzynski, Margaret L. Kern, Eduardo Blanco, Michal Kosinski, David Stillwell, Martin EP Seligman, and Lyle H. Ungar. "Toward personality insights from language exploration in social media." In 2013 AAAI Spring Symposium Series. 2013.

[8] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." Ain Shams engineering journal 5, no. 4 (2014): 1093-1113.

[9] Mäntylä, Mika V., Daniel Graziotin, and Miikka Kuutila. "The evolution of sentiment analysis—A review of research topics, venues, and top cited papers." Computer Science Review 27 (2018): 16-32.

[10] Khan MN, Jamil M, Gilani SO, Ahmad I, Uzair M, Omer H. Photo detector-based indoor positioning systems variants: A new look. Computers & Electrical Engineering. 2020 May 1;83:106607.

[11] Kashif H, Khan MN, Altalbe A. Hybrid Optical-Radio Transmission System Link Quality: Link Budget Analysis. IEEE Access. 2020 Mar 18;8:65983-92.

[12] Zafar K, Gilani SO, Waris A, Ahmed A, Jamil M, Khan MN, Sohail Kashif A. Skin Lesion Segmentation from Dermoscopic Images Using Convolutional Neural Network. Sensors. 2020 Jan;20(6):1601.

[13] Uzair M, D DONY RO, Jamil M, MAHMOOD KB, Khan MN. A no-reference framework for evaluating video quality streamed through wireless network. Turkish Journal of Electrical Engineering & Computer Sciences. 2019 Sep 18;27(5):3383-99.

[14] Talib, Ramzan, Muhammad Kashif Hanif, Shaeela Ayesha, and Fakeeha Fatima. "Text mining: techniques, applications and issues." International Journal of Advanced Computer Science and Applications 7, no. 11 (2016): 414-418.

[15] Öztürk, Nazan, and Serkan Ayvaz. "Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis." Telematics and Informatics 35, no. 1 (2018): 136-147.

[16] Ptaszynski M, Rzepka R, Araki K, Momouchi Y. Research on emoticons: review of the field and proposal of research framework. Proceedings of 17th Association for Natural Language Processing. 2011 Mar 7:1159-62.

[17] Perveen N, Missen MM, Rasool Q, Akhtar N. Sentiment based twitter spam detection. International Journal of Advanced Computer Science and Applications (IJACSA). 2016 Jul 1;7(7):568-73.