

Twitter Sentimental Analysis

Shobana G, Vigneshwara B, Maniraj Sai A.

Abstract: In this current era, social media plays a important role in data exchange, sharing their thoughts. Emotional Effect of a person maintains an important role on their day to day life. Sentiment Analysis is a procedure of analyzing the opinions and polarity of thoughts of the person. Twitter is a main platform on sharing the thought's, opinion and sentiments on different occasions. Twitter Sentimental Analysis is method of analyzing the emotions from tweets (message posted by user in twitter). Tweets are helpful in extracting the Sentimental values from the user. The data provide the Polarity indication like positive, negative or unbiased values. It is focused on the person's tweets and the hash tags for understanding the situations in each aspect of the criteria. The paper is to analyse the famous person's id's (@realdonaldtrump) or hash tags (#IPL2018) for understanding the mindset of people in each situation when the person has tweeted or has acted upon some incidents. The proposed system is to analyze the sentiment of the people using python, twitter API, Text Blob (Library for processing text). As the results it helps to analysis the post with a better accuracy.

Keywords: (@realdonaldtrump), (#IPL2018), Text Blob (Library for Processing Text).

I. INTRODUCTION

In the past years, the young generation people are moving towards the social media like Google Plus, WhatsApp, Facebook, Twitter, etc. The social media is also revolving with those people to get them involved by making current trending insights concepts that is trending within a second. In the recent years, the people are exposing their social related issues through several social media by comments, reviews, posts, hashtags, emoji's, etc. which was followed by many people and those tweets become popular soon. Moreover, the social media is also bringing tremendous opportunity platform for businesses to connect with the consumers so easily. People rest on mostly user produced content like, comments, over online for making the decision. Example: if anyone has to buy a product or make a decision, they initially search its reviews online, converse about it on social media. The content that is displayed for that product is mainly taken into the point as well as the discussion in the social media is also noticed and these made the way to make our business a success. To automate our analysis based on the reviews or comments in the social media by the people, for a sentimental analysis. Sentimental Analysis (SA) is introduced to the world to tell us the information is correct or wrong in each scenario using the social media tags.

Manuscript published on 30 November 2018.

*Correspondence Author(s)

Shobana G, Assistant Professor, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

Vigneshwara B, UG Scholars, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

Maniraj Sai A, UG Scholars, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore (Tamil Nadu), India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Thus, we can know about how world or people are reacting to every aspect currently going in the world.

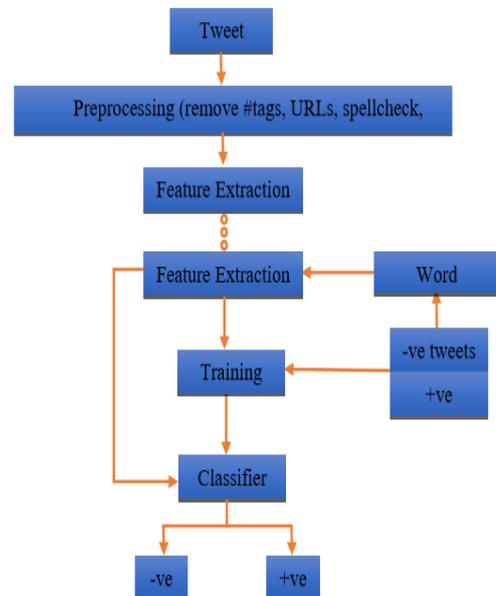


Fig 1.1 System Architecture

The system architecture consists of the components as shown in the figure 1.1 such as Tweets extraction from twitter, preprocessing of data, feature extraction, Training set are defined for the given analysis. The training set is obtained by predefined set of positive or negative tweets which can be done using naive Bayes or support vector machine (SVM) and output obtained is positive, negative tweets. The Classifier will classify the tweets according to the training set and regulates the polarity of the tweet as the output. In this paper, we are going to analyze the microblog called as Twitter, classify the "tweets" into positive, negative and neutral sentiment. We explore the method for building such data using Twitter hashtags (e.g., #best feeling, #DonaldTrump, #love) to identify positive, negative, and neutral tweets to use for training three-way sentiment classifiers. Thus, these tweets and the hashtags are must for analyzing the thinking level of individual people.

II. LITERATURE SURVEY

Sentiment analysis is the process of analysis of the text from many levels. First level is document level [3], the classification task determine the class of an object based on its attributes (Turney, 2002; Pang and Lee, 2004), and after that it can be analysed at the sentence level [5] for classifying the sentence based on the negative, positive and neutral sentiments (Hu and Liu, 2004; Kim and Hovy, 2004) and next level is the phrase level [4]



for defining if an expression is unbiased or polar and then remove uncertainty of meaning from the polarity of the polar expressions (Wilson et al., 2005; Agarwal et al., 2009. Bermingham and Smeaton(2010) and Pak and Paroubek (2010). Go et al. (2009) they used distant learning algorithm to obtain the sentiment data [8]. In this techniques, positive emoticons symbols in tweets such as “:)” “:-)” and negative emoticons symbols in tweets such as like “:(” “:-(”. They proposed the models using Naive Bayes algorithm for analysis the text and the report are generated and visualized.

They used two methods such as unigrams for identifying single word repeating over the context and bigrams for identifying double word repeating over the context along with Parts-of-Speech (POS) for analysing the tweets. But the unigram method had reached a better way of analysis but the bigrams and POS had failed to attempt his purpose.

Pak and Paroubek (2010) [2] collect the following tweets considered as data which really helped them in similar distant learning paradigm for setting a model for analysis. They perform classification of task such as subjective, objective. For subjective the information are get from the user tweets by means of text or image or symbols as Go et al. (2009) [8]. For objective information the information are obtained from verification of the data such as famous newspapers like “Times of India”, “Washington Posts” etc. Information which is taken for analysis is casual sample of flowing tweets collected by using queries. In the past year there have been numerous documents observing the Twitter sentiment and buzz [1], [2], [4] (Jansen et al. 2009; Pak and Paroubek 2010; O’Connor et al. 2010; Tumasjan et al. 2010; Bifet and Frank 2010; Barbosa and Feng 2010; Davidov, Tsur, and Rappoport 2010). Further scientists have started to discover the usage of part-of-speech structures but results remain mixed. It has enormous interesting chances to develop the innovative applications, because success of many business depends on accessible information on online sources such as blogs, twitter and other social networks.

Barbosa and Feng (2010)[4] has analysed the sentimental classification on Twitter data. The test data of tweets are collected, they have taken some of the syntax features for analysis of tweets which contains symbols, retweet, emoticons, tags, link, punctuation and exclamation marks, semicolon are in the combination with structures for identifying the polarity of words.

Kamps et al. (2002) [12] has analysed the data by using the lexical database. Lexical database is description of lexemes. Lexical database such as WordNet are used. This contains the emotional content of a word. The distance metric of words are used to determined semantic polarity of adjectives.

Researchers are also trying to find different ways of analysing tweets based on the ideas they had while understanding the concept. Researchers tried this analysis using some of the specified fields such as Machine learning which uses Naive Bayes, Maximum entropy and SVM alongside the Semantic Orientation based Word Net which extracts equivalent words and similitude for the content feature, then Lexicon based analysis based on the created dataset which consists of pre-processed tweets and lastly, Hybrid approach where some researchers combined the

supervised machine learning and lexicon based approaches together to improve sentiment classification performance. Gamon (2004) [9] has done sentiment analysis on feedback data from the Global Support Services survey. They are used query to identify the role of features like Part of Speech tags. The accuracy of classifier can be obtained by some of factors such as feature selection, from the testing data and demonstrate the abstract linguistic analysis feature for accuracy of data.

Devaki.p, et al (2017)[15] has done analysis on twitter data for election. It indicates the popularity of parties in the election based on positive tweets. This system uses Naive Bayes classifier algorithm are used to classify the positive and negative tweets.

A comparative study of existing techniques for mining the data which includes machine learning, Interdependent Latent Dirichlet Allocation, lexicon-based approaches, together with cross domain, cross-lingual methods and some evaluation metrics. The concept level sentences analysis uses the Combining Lexicon and Learning based Approach. As the result of study, machine learning methods such as Support Vector Machine and Naive Bayes have the highest accuracy and can be regarded as the baseline learning methods, while lexicon-based methods are very effective in some cases.

More research is needed to determine whether the POS features are just of poor quality due to the results of the tagger or whether POS features are just less useful for sentiment analysis in this domain. Features from an existing sentiment lexicon were somewhat useful in conjunction with microblogging features, but the microblogging features (i.e., the presence of intensifiers and positive/negative/neutral emoticons and abbreviations) were clearly the most useful. In this paper, we perform extensive feature analysis of tweets using hashtags, ID’s and building model classifications.

III. METHODOLOGY

In this method we uses textblob as a method to find the polarity of the text (positive text, negative text or neutral text). The tweets are imported from the Twitter using the (API) provided by the Twitter Developer. From these API various fields like tweets, source, retweets, likes, language, user etc. can be scrapped. After collecting these data, we can analyses the various famous person thoughts on an event or occasion

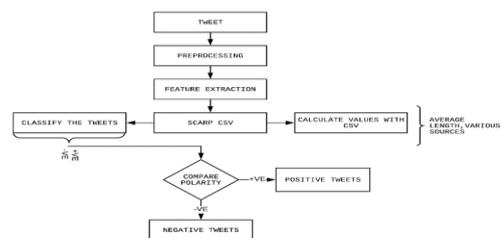


Fig 3.1 Architectural Flow of Twitter Analysis



The figure 3.1 explains the extraction of tweets id from twitter through API, then preprocess the data that are extracted. Preprocessing includes exclusion of unwanted fields, segregating the fields important for analysis. Once the fields are extracted and segregated CSV is created. Using this CSV, the length of the message, Likes, retweets for the id is extracted and various results are derived. With the scraped tweets, classify the tweets whether positive or negative or neutral.

A. Dataset Description

In this proposed system, we have used the dataset called astwitterdataset. csv. It contains the following fields Tweets, Len, ID, Date, Source, Likes, RT's (Retweets), SA (Sentimental Analysis).

B. Software Description

In the system the graphs such as Table, Bar graph, Line graph are generated with the help of Spyder and Jupyter notebook. The predefined functions are pandas, numpy, matplotlib, pyplot, list, Dictionary. Pandas is used for converting from csv file to dataset. Numpy is one of the essential library for scientific calculating in Python. It delivers a high-performance multidimensional array object, and apparatuses for experementing with these arrays. Python comprises of numerous built-in container categories: lists, dictionaries, sets, and tuples. A list is the Python equal of an array, but is resizable and can contain elements of different types. A dictionary stores (key, value) pairs, like a Map in Java or an object in JavaScript. Python library such as Text Blobare used for processing the textual data. It provides API for processing natural language processing (NLP) such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more. Tweepyisused for accessing Twitter API and it is open sourced.

C. Data Analysis and Visualization

In Twitter the various famous personalities tweet their thoughts on their opinion on an occasion. From their thoughts, importance of that occasion and the polarity of their tweet are analysed. Some of the analysis with the dataset as follows.

- Calculate the average length of the tweet and visualize the average length for a period.
- Visualize the favorites and retweets for each personality.
- Visualize the various source of the tweet.
- Calculate the polarity of the tweets given by the person
- Visualize the Polarity of tweet (positive, negative, neutral)
- Compare the tweets polarity of various famous personalities

Analyses to calculate the average length of the tweet and visualize the average length for a period.

Calculate the average length of the tweet of the person using the mean function and visualize the average length for a period. With the calculated average. Visualize the average length of tweet for a period.

The length's average in tweets: 130.415

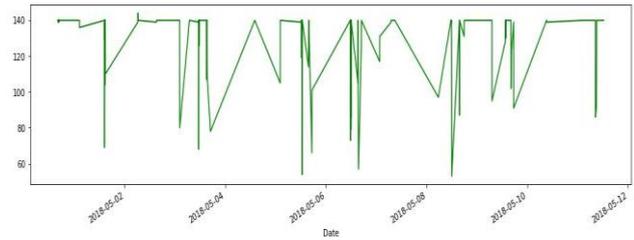


Fig 3.2: Average length of Tweets for a given period
Analysistovisualize the favorites and retweets for each personality.

For Visualizing the favorites and retweets for each personality, it is needed to calculate the number of Favorites and RT's from CSV. Visualize the data and compare with various personalities.

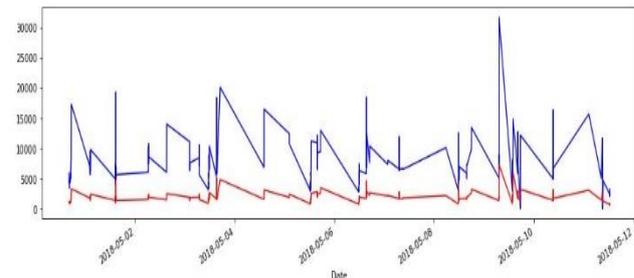


Fig 3.3: Narendra Modi Favorites and retweets

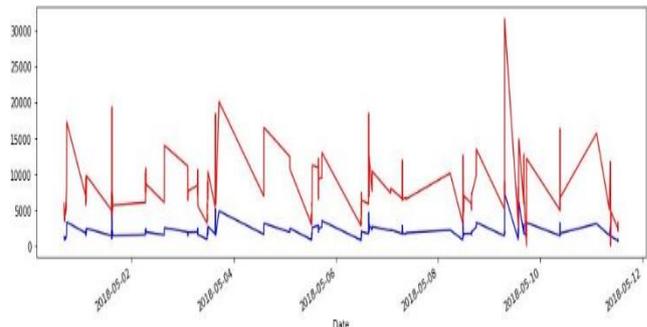


Fig 3.4 Rahul Gandhi Favorites and retweets
Analysis to visualize the various source of the tweet.

Twitter tweets are tweeted through various sources like Media Source, iPhone or android phone. Visualize the number of sources used by a user to tweet his opinion in twitter. Most used source for tweets by the personalities can be identified.

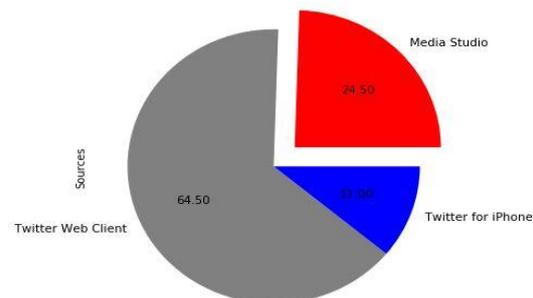


Fig 3. 5. Source for Narendra Modi Tweets

IV. CONCLUSION

Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model to train it and then check its accuracy, so that we can use this model for future use according to the results. It comprises of steps like data collection, text pre-processing, sentiment detection, sentiment classification, training and testing the model. This research topic has evolved during the last decade with models reaching the efficiency of almost 85%-90%. But it still lacks the dimension of diversity in the data. Along with this it has a lot of application issues with the slang used and the short forms of words. Many analyzers don't perform well when the number of classes are increased. Also, it's still not tested that how accurate the model will be for topics other than the one in consideration. Hence sentiment analysis has a very bright scope of development in future.

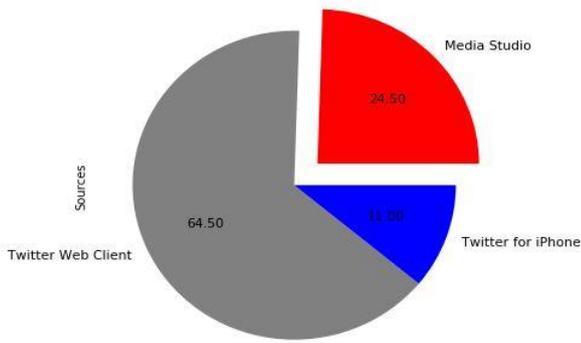


Fig 3.6 Sources for Rahul Gandhi Tweets

Analysis to calculate the polarity of the tweets given by the person

Using Textblob calculate the polarity of each tweets and calculate the average polarity level of the tweets and interfere that the person tweets polarity level. This allows to find the polarity of the tweet on an occasion. Calculate and store the three levels in the list.

Analysis to visualize the polarity percent

From the calculated polarity we can visualize the polarity through charts. Compare the polarity with various personalities and conclude it with the personalities who tweets maximum positive tweets.

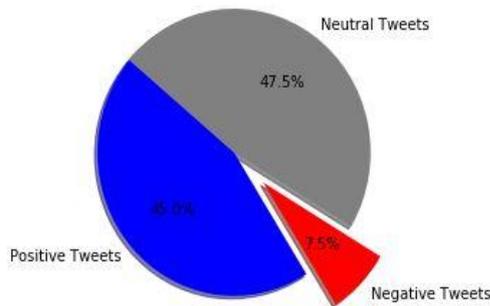


Fig 3.7 Polarity level of Narendra Modi tweet.

Comparison of the tweets polarity of various famous personalities

From the calculated polarity levels of various personalities, we can compare their polarity level to find the polarity of their tweet, whether their tweet is positive or negative or neutral. Visualize various personalities with their tweets polarity and compare.

COMPARISON BETWEEN SENTIMENTAL ANALYSIS OF TWO LEADERS

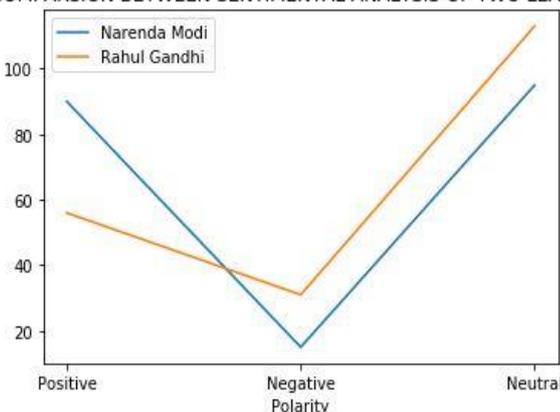


Fig 3.8 Comparison of the Polarity Levels

REFERENCES

- Jansen,B.J.; Zhang,M.; Sobel,K.; and Chowdury,A. (2009), "Twitterpower: Tweets as electronic word of mouth", Journal of the American Society for Information Science and Technology 60(11):2169-2188.
- Pak, A., and Paroubek, P (2010), "Twitter as a corpus for sentiment analysis and opinion mining". In Proc. of LREC.
- Pang, B., and Lee, L. (2008), "Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval" 2(1-2):1-135.
- Wilson, T. Wiebe, J.; and Hoffmann, (P. 2009),"Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Linguistics", 35(3):399-433.
- M Hu and B Liu. (2004),"Mining and summarizing customer reviews. KDD".
- L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitter from Biased and Noisy Data". COLING 2010: Poster Volume, pp. 36-44.
- J. Kamps, M. Marx, R. J. Mokken, and M. De Rijke. "Using wordnet to measure semantic orientations of adjectives," 2004.
- Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. Technical report, Stanford.
- David Zimbra, M. Ghiassi and Sean Lee, "Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", IEEE 1530-1605, 2016.
- Varsha Sahayak, Vijaya Shete and Apashabi Pathan, "Sentiment Analysis on Twitter Data", (IJIRAE) ISSN: 2349-2163, January 2015.
- Peiman Barnaghi, John G. Breslin and Parsa Ghaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", 2016 IEEE Second International Conference on Big Data Computing Service and Applications.
- Mondher Bouazizi and Tomoaki Ohtsuki, "Sentiment Analysis: from Binary to Multi-Class Classification", IEEE ICC 2016 SAC Social Networking, ISBN 978-1-4799-6664-6.
- Nehal Mangain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt, "Sentiment Analysis of Top Colleges in India Using Twitter Data", (IEEE) ISBN -978-1-5090-0082-1, 2016.
- [https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python.](https://www.geeksforgeeks.org/twitter-sentiment-analysis-using-python/) "Twitter Sentimental analysis for Realdonaldtrump" Devaki P, Ilakiya, J, Indumathi, R and Arul Priya, M,
- "Geospatially and literally analysing tweets", Journal of Advanced Research in Dynamical and Control Systems Volume 9, Issue Special Issue 14, 2017, Pages 1002-1009