IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Special Issue, 4, Feb 2019 E-ISSN: 2347-2693

Emotion Identification in Tweets Using NLP And Classification Procedure

N. Vasunthira Devi[1*], R. Ponnusamy[2]

[1]Department of Computer Applications,Annai Vailankanni Arts And Science College, Thanjavur, Tamilnadu.
[2]Department of Computer Science & Engineering,CVR College of Engineering, Hyderabad.

Corresponding Author: vasunthira@gmail.com

Available online at: www.ijcseonline.org

Abstract—This paper mainly focuses on the classification of the tweets based on the emotion in which they prompt. The proposed method will extract the tweets on any particular issue and will assist us to analyze the opinion of the people. The tweets can be classified as positive negative or neutral against the particular issue in which the query was made to extract the tweets. The technology which we use is the twitter API, which will assist us in extracting the tweets relevant the particular issue. The next step is to process the tweets i.e. here we will remove all unwanted images, punctuations and special characters. And at last all the tweets will be converted into lowercase for further steps. The classification of final processed tweets will employ a supervised classification method. The basic classifier used in this method is Naive Bayes classifier to classify the emotion of the tweets. The algorithm is trained by all the possible extent. Finally the percentage of the positive and negative tweets will be calculated. Based on the graphical representation we can create a new strategy for the particular issue.

Keywords— Tweets, Supervised Classification, Positive and Negative tweet.

I. INTRODUCTION

Twitter is a mainstream microblogging administration where clients make status messages (called "tweets"). These tweets once in a while express feelings about various points. We propose a technique to consequently remove emotion (positive or neutral or negative) from a tweet. This is helpful on the grounds that it enables criticism to be accumulated without manual intercession. Buyers can utilize emotion examination to look into items or administrations before making a buy. Advertisers can utilize this to explore general supposition of their organization and items or to investigate consumer loyalty. Associations can likewise utilize this to assemble basic criticism about issues in recently discharged items. There has been a lot of research in the zone of emotion classification. Generally, a large portion of it has concentrated on grouping bigger bits of content, similar to reviews. Tweets (and microblogs as a rule) are not the same as reviews basically due to their motivation: while reviews speak to abridged contemplations of creators, tweets are progressively easygoing and constrained to 140 characters of content. For the most part, tweets are not as attentively made as reviews. However, despite everything they offer organizations an extra road to accumulate input. Past research on investigating blog entries by PANG ET AL. have examined the execution of various classifiers on film reviews. Crafted by Pang et al. has filled in as a standard and numerous creators have utilized the strategies given in their work crosswise over various spaces. So as to prepare a classifier, administered adapting, for the most part, requires hand-labelled preparing information. With the expansive scope of subjects talked about on Twitter, it would be exceptionally hard to physically gather enough information to prepare a emotion identificationfor tweets. Subsequently, we have utilized openly accessible twitter datasets. Be that as it may, this dataset comprises just positive and negative tweets. For neutral tweets, we have utilized the freely accessible neutral tweet dataset gave. We run the machine learning classifiers Naïve Bayes prepared on the positive and negative tweets dataset and the neutral tweets against a test set of tweets. This can be utilized by people and organizations that might need to inquire about emotion on any subject.

II. BACKGROUND

A.Defining the Emotion

For the projected work the proposed algorithm label's all review to positive or negative based on the polarity of the review text. All the review which are not labeled are considered as neutral

Example for Emotion Identification of Tweets
Positive: the climate is very good to enjoy.
Negative: I hate this place.
Neutral: He comes to office at 9 in the morning.

III. METHODOLOGY

A. Pre-processing

1.The Twitter language demonstrate has numerous exceptional properties. These properties can be utilized to decrease the element space:

2.UsernamesSo as to coordinate their messages clients frequently incorporate twitter usernames in their tweets. A true standard is to incorporate @ image before the username (for example @towardshumanity). A class token (AT_USER) replaces all words that start with @ image.

3. Usages of links:Clients frequently incorporate connections in their tweets. To disentangle our further work, we convert a URL like "http://tinyurl.com/cmn99f" to the token "URL".

4. Stop words:There are a great deal of stop words or filler words, for example, "an", "is", "the" utilized in a tweet which does not demonstrate any emotion and henceforth these are sifted through.

5. Repeated letters:Tweets contain easygoing language. For instance, in the event that you seek "hi" with a discretionary number of „i‟s in the center (for example hiiiiiiiiiii) on Twitter, there will undoubtedly be a nonempty result set. I use pre-preparing so any letter happening multiple occasions in succession is supplanted with two events. In the examples over, these words would be changed over into the token "hi".

B. Feature Vector

After pre-preparing the tweets, we get highlights which have square with loads. Unigram Features which are independently enough to comprehend the emotion of a tweet is called as unigram. For instance, words like „good‟, happy‟ obviously express a positive emotion.

C. Classification

With the end goal of classification of tweets, we make use of Naïve Bayes classifier. Naïve Bayes is a probabilistic classifier dependent on Bayes‟ hypothesis. It orders the tweets dependent on the likelihood that a given tweet has a place with a specific class. We consider three classes to be specific, positive, negative and neutral. We appoint class c* to tweet d where,

$$c* = argmac_c P_{NB}(c|d)$$

$$P_{NB}(c|d) := \frac{(P(c)\sum_{i=1}^{m} P(f|c)^{n_i(d)})}{P(d)}$$

In this equation, f speaks to a component and ni(d) speaks to the inclusion of highlight fi discovered tweet d. There is an aggregate of m highlights. Parameters P(c) and P(f|c) are acquired through most extreme probability gauges and smoothing is used for inconspicuous highlights. We have utilized the Python based Natural Language Toolkit library to prepare and characterize utilizing the Naïve Bayes technique.

IV. EXPERIMENTAL RESULTS

There are freely accessible informational indexes of Twitter messages with emotion shown by. We have utilized a mix of these two datasets to prepare the machine learning classifiers. For the test dataset, we haphazardly pick 4000 tweets which were not used to prepare the classifier. The Twitter API has a parameter that determines which language to recover tweets in. We constantly set this parameter to thanglish (en). In this manner, our classification will just work on tweets in Thanglish on the grounds that the preparation information is thanglish-as it was. We assemble a web interface which scans the Twitter API for a given catchphrase for as far back as one day or seven days and gets those outcomes which are then exposed to preprocessing. These separated tweets are nourished into the prepared classifiers and the subsequent yield then appears as a chart in the web interface.

V. CONCLUSION

A live Twitter channel is gathered under the catchphrases entered by the client. The feed is put away in a MongoDB database. It is likewise put away locally in a json document. The information was pre-prepared to evacuate pointless spaces, images and futile highlights. Despite everything it requires further work to evacuate however much clamor as could reasonably be expected. Roughly more than 2000 tweets are then put away as a csv document for examination. Various Lexicon put together techniques are used with respect to singular tweets from the record to evaluate their value. The picked classifier for this work is a Naive Bayes Classifier using the content handling apparatuses in NLTK and their ability to work with human language information. It is prepared on labeled tweets and afterward used to break down the emotion in the tweets about the looked point. The outcome is spoken to as a pie graph which demonstrates the level of clients who have positive supposition on the looked point when contrasted with the ones have negative conclusion or are neutral.

VI. FUTURE WORK

Machine learning methods perform well to arrange emotion in tweets. We trust the precision of the framework could be as yet made strides. The following is a rundown of thoughts we think could encourage the classification: 1.Semantics:The extremity of a tweet may rely upon the point of view you are deciphering the tweet from. For instance, in the tweet "Federer beats Nadal :)", the emotion is positive for Federer and negative for Nadal. For this situation, semantics may help. Utilizing a semantic job labeler may demonstrate which thing is essentially connected with the action word and the classification would happen as needs be. This may permit "Nadal beats Federer :)" to be characterized uniquely in contrast to "Federer beats Nadal :)".2. Internationalization:

Presently, we center just aroundthanglish tweets yet Twitter has an enormous global crowd. It should be conceivable to utilize our way to deal with arrange emotion in different languages with a language explicit positive/negative watchword list.

REFERENCES

[1] V. Kagan, A. Stevens, V. S. Subrahamian, "Using Twitter sentiment to forecast the 2013 Pakistani Election and the 2014 Indian Election", Journal of IEEE Intelligent Systems, vol. 30, pp. 2-5, 2015.

[2] Min Song, MeenChul Kim, Yoo Kyung Jeong, "Analyzing the Political Landscape of 2012 Korean Presidential Election in Twitter", IEEE Computer Society,[online],Available: www.computer.org/intelligent.

[3] Shenghua Liu, Xueqi Cheng, Fuxin Li, Fangtao Li, "TASC:Topic Adaptive Sentiment Classification on Dynamic Tweets", IEEE transactions on knowledge and data engineering, vol. 27, no. 6, pp. 1696-1709, june 2015.

[4] Mark E. Larsen, Tjeerd W. Boonstra, Philip J. Batterham, Bridianne O'Dea, Cecile Paris, Helen Christensen, "We Feel: Mapping Emotion on Twitter", IEEE journal of biomedical and health informatics, vol. 19, no. 4, pp. 1246-1252, july 2015.

[5] Rui Xia, Feng Xu, ChengqingZong, Qianmu Li, Yong Qi, Tao Li, "Dual Sentiment Analysis on tweets: Considering Two Sides of One Review", IEEE transactions on knowledge and data engineering, vol. 27, no. 8, pp. 2120-2133, august 2015.

[6] Shulong Tan, Yang Li, Huan Sun, Ziyu Guan, Xifeng Yan, Jiajun Bu, Chun Chen, Xiaofei He, “Interpreting the Public Sentiment Variations on Twitter”, IEEE Transactions on Knowledge and Data Engineering, vol. 26, pp. 1158-1170, 2014.

[7] HaseSudeepKisan, HaseAnandKisan, AherPriyanka Suresh, “Collective intelligence & sentimental analysis of twitter data by using Standford NLP libraries with software as a service (SaaS)”, IEEE International Conference on Computational Intelligence and Computing Research, pp. 1-4, 2016.

[8] HaseSudeepKisan, HaseAnandKisan, AherPriyanka Suresh, “Sentiment Analysis on Twitter Using Streaming API”, IEEE International Advance Computing Conference (IACC), pp. 915-919, 2017.

[9] MestanFıratÇeliktuğ, “Twitter Sentiment Analysis, 3-Way Classification: Positive, Negative or Neutral”, IEEE International Conference on Big Data (Big Data), pp.2098- 2103, 2018.

[10] Saki Kitaoka,TakashiHasuike, “Where is safe: Analyzing the relationship between the area and emotion using Twitter data”, IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1-8, 2017.

[11] M. Muralidharan,Dr. J. Jeyachidra, "SCARTA: Selection Conscious Approach to Retrieval ofTherapeutic Archives", International Journal of Pure and Applied Mathematics,Volume 119 No. 153461-3469, 2018.

[12] WanxiangChe, Yanyan Zhao, HongleiGuo, Zhong Su, Ting Liu, "Sentence Compression for Aspect-Based Sentiment Analysis", IEEE/ACM transactions on audio speech and language processing, vol. 23, no. 12, pp. 2111-2124, December 2015.

[13] Nan Cao, Conglei Shi, Sabrina Lin, Jie Lu, Yu-Ru Lin, Ching-Yung Lin, "TargetVue: Visual Analysis of Anomalous User Behaviors in Online Communication Systems", IEEE transactions on visualization and computer graphics, vol. 22, no. 1, pp. 280-289, january 2016.

[14] Erik Cambria, BjörnSchuller, Yunqing Xia, Catherine Havasi, "New Avenues in Opinion Mining and Sentiment Analysis", IEEE Computer Society, March-April 2013.

[15] Xin Chen, MihaelaVorvoreanu, Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences", IEEE transactions on learning technologies, vol. 7, no. 3, pp. 246-258, july-september 2014.

[16] Desheng Dash Wu, LijuanZheng, David L. Olson, "A Decision Support Approach for Online Stock Forum Sentiment Analysis", IEEE transactions on systems man and cybernetics: systems, vol. 44, no. 8, pp. 1077-1084, august 2014.

[17] M. Muralidharan,V. ValliMayil, "A Study of Natural Language Processing Procedures", IJCSE E-ISSN: 2347-2693, vol. 5, 2017.

