

Predicting levels of depression from social network data using machine learning technique

Siddhesh Khanvilkar, Suhani Ture, Bhavya Shah, Nikhil Zade

Abstract— One of the biggest differences in the lives of current teenagers and young adults, as compared to earlier generations, is that they spend much less time connecting directly with peoples and more time connecting electronically, principally through social media. Some experts see the rise in depression as evidence because social media users are less emotionally satisfied. Twitter become the most popular social media platform that allow people to share the information through small messages called tweets on a real time basis. The proposed method uses concepts like Natural Language Processing for text analysis. This system will analyze the sentiments as positive or negative using Textblob's sentiment method based on that the system will detects the different levels of depression.

Index Terms— Depression, Emotions, Social Media, Twitter, Machine Learning, Sentiment Analysis, NLP

1 INTRODUCTION

The communication technologies, specially the online social networks have rejuvenated how people interact and communicate with each other through electronic media. The applications such as Twitter, Facebook, Instagram and alike not only host the written and multimedia contents but also offer their users to express their feelings, emotions and sentiments about a topic, subject or an issue online. This is great for users of social networking site to openly and freely contribute and respond to any topic online; on the other hand, it creates opportunities for people working in the health sector to get insight of what might be happening at mental state of some-one who reacted to a topic in a specific manner.

To provide such insight, machine learning techniques potentially offer some unique features that can assist in examining the unique patterns hidden in online communication and process them to reveal the mental state such as 'happiness', 'sadness', 'anger', 'anxiety', depression among social networks' users. Moreover, there is growing body of literature addressing the role of social networks on the structure of social relationships such as breakup relationship, mental illness such 'depression', 'anxiety', 'bipolar' etc., smoking and drinking relapse, and sexual harassment.

2 Methodology

In this study , we focused on tweepy library which is used to fetched Real-time data of Twitter .The inbuilt tweepy API provides access to the entire twitter RESTful API methods. Each method can accept various parameters and return responses.

The inbuilt API used to access the tweet and analyse them. After fetching the tweets ,perform sentimental analysis on each of the tweets using textblob and based on the polrity value, we obtain the result that particular tweet is positive or negative. we apply unsupervised machine learning approach to study each factor independently. Based the analysis of result pie chart is generated.

3 Workflow Of Sentiment Analysis

3.1 Procedure of Sentiment Analysis:

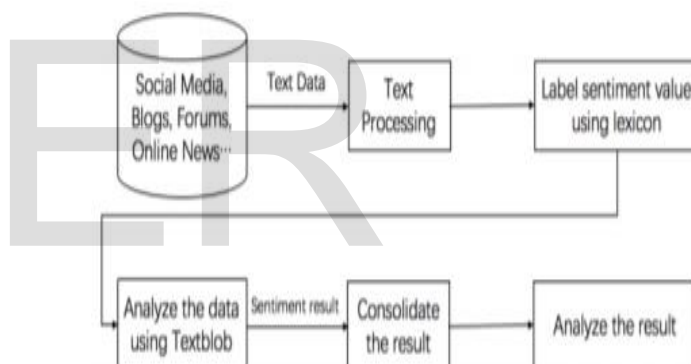


Fig 1- basic framework of predicting the levels of Depression

According to Figure 1, the process of predicting depression includes the following steps: First, get the previous tweets data from twitter and preprocess the text data. In this step, the data will be cleaned and simplified. Tokenization is also included in this step to split a stream of text into smaller units, which is helpful to analyse the data from a statistical point of view. Second, extract the necessary data from the pre-processed text to satisfy the requirement of analysis, and build and improve the lexicon of the tool. Finally, will get the pie-chart of sentiments. Based on that, we can predict the depression level.

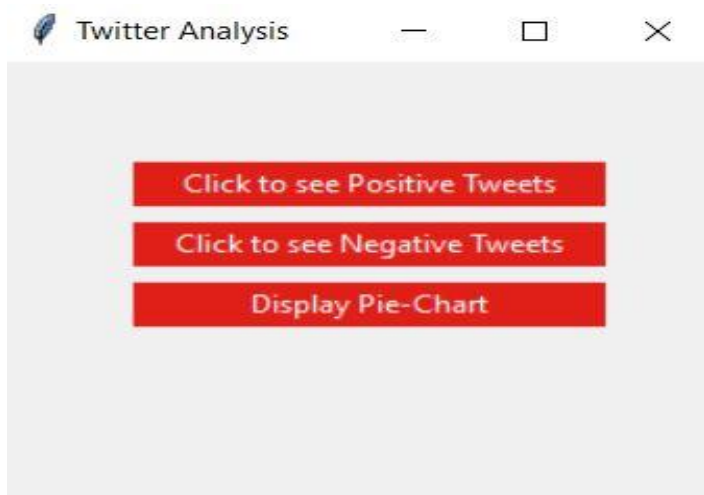


Fig 2- Twitter analysis

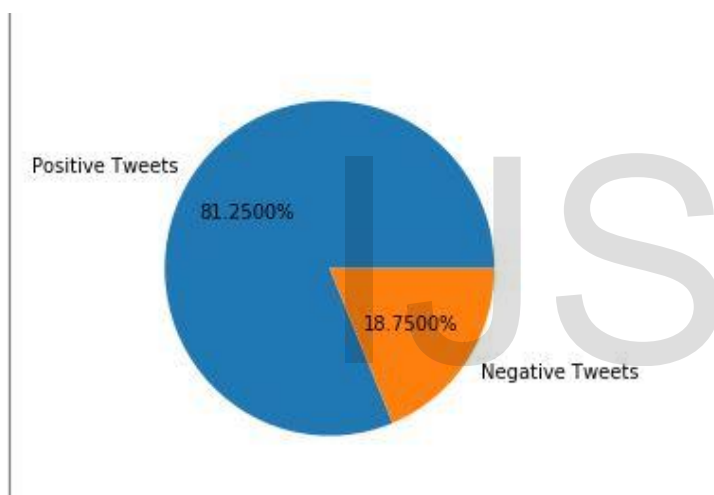


Fig 3- Pie-chart of Negative and Positive Tweets

3.2 Data Mining and Text Processing:

Data mining is a step towards data availability. Usually reference to a large number of data using algorithm search to find the information hidden in the process. It has long been used in marketing, health, engineering, and information systems and is usually related to computer science. In statistics, analysis and online work, details retrieval, machine learning, professional programming and pattern recognition and many other methods, the above objectives may apply be achieved. Twitter provides a lot of API for developers to have low latency access to tweet information. For example, the streaming API and the rest API are widely used APIs. The tweet search API can only search for the latest tweets published in the last 7 days, while the streaming API can download real-time tweets. For extraction of tweets we had used tweepy library. Tweepy supports both OAuth 1a (application-user) and OAuth 2 (ap-

plication-only) authentication. Authentication is handled by the tweepy.AuthHandler class. Tweepy tries to make OAuth 1a as painless as possible for you. To begin the process we need to register our client application with Twitter. Create a new application and once you are done you should have your consumer key and secret. Keep these two handy, we'll need them.

Generally, text mining involves the following steps: 1) data acquisition and text correction, 2) feature extraction and weight distribution, 3) model development and training in the model. Text information needs to be reproduced after sufficient text analysis has been obtained. The main steps in writing text encoding are word classification, deletion, and other re-downloads. Classification technology used in this paper is a token function in the NLTK library and a classification namespace package in Textblob.

The former is used to separate every detail and calculate the frequency of each word, and the last one is used to divide a piece of twitter into words and determine the amount of sound. With the help of these tools, the process of tokenization can be completed with high accuracy. In addition, these packages also support the addition of syntactic dictionaries, enabling us to add more emotional words to the subject dictionary, which enables some judgmental results to meet the requirements of subsequent follow-up analyzes. After the text is marked, the fixed-word words are subtracted from the text amplification algorithm to include the steps for the removal of the emotional words and the emotional element of the word polarity judgment.

3.3 Sentiment Analysis:

Sentiment analysis is to analyze the collection of personal choices, also known as emotions mining and ideas. Depending on the size of the text, Sentiment analysis can be divided into three categories: 1) Emotional analysis based on words. It creates feelings lexicon to distinguish good words from bad words or automatically, and then uses a dictionary to analyze the information based on the frequency of the word or the amount of sensitivity which is set when a dictionary is created. 2) Sensitivity analysis based on sentences. It analyzes in particular the relevant aspects of a tendency such as a semantic statement and sentence sentences and so on, combined with unity of change and bad words. 3) Emotional analysis based on document. Basically, of course much like the second paragraph, because the statement can be it is seen as a short text. In this paper, a sensitivity analysis it involves different techniques such as the invention of subject dictionary, word-based analysis and a label for the amount of sensitivity that is positive, negative or neutral.

A word analysis is the basis of an analysis of the feelings of the text, and the nature of the sentences and the sentiment analysis of the chapters. Lots of feelings Investigators focus on the emotional divide of the text; too it is rare to write the power of

synonyms. This is the type of distinction sometimes only separates words into positive, negative and neutral categories, and they are not marked the right kinds of emotions and the power of emotions. The emotional unity of the default text is normal it is divided into two poles: constructive informative and a guarantee, and a negative one for criticism and apathy. The ability to understand the text to find the power of a low spirit, such as a strong, regular blasphemy contempt, purpose, general praise, and strong praise.

No matter, whichever classification tool is used, how to separate the good or bad words that is always shown in the lexicon. In this study, Textblob is used to combine the two types of judgment together and separate positive tweet and negative tweet, and give the result of analysis sentiment analysis. Araque (2017) revealed that TextBlob is more suited to the Twitter domain than other divisors. It an open source text library written in Python, which provides a consistent API for sinking into a standard NLP functions such as partial speech marking, pronunciation, Nervous analysis, and so on..

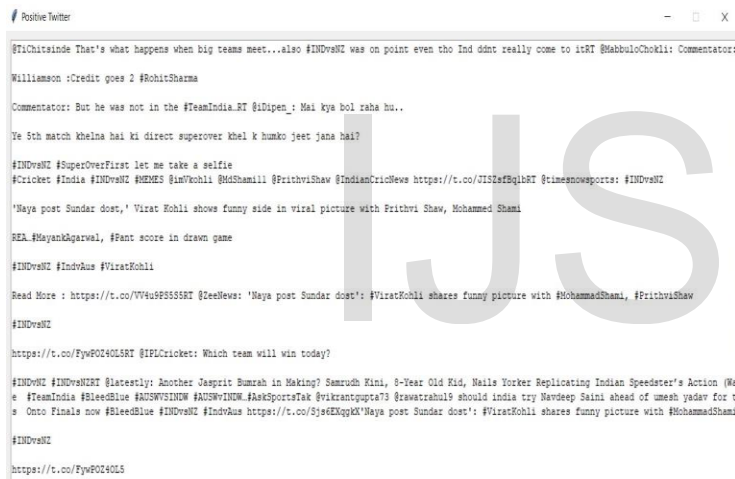


Fig 4- Negative Tweets



Fig 4- Positive Tweets

4 Conclusion

We have presented the system that take real-time data from twitter and perform sentiment analysis to detect the different levels of depression. The information of different levels of depression are displayed in the form of percentage and pie chart. In this paper we used Twitter for monitoring and detecting the depression among its users.

Acknowledgment

I would like to express my special thanks of gratitude to my guide (Siddhesh Khanvilkar) who gave me the golden opportunity to do this wonderful project on the topic (Predicting levels of depression from social network data using machine learning techniques), which also helped me in doing a lot of Research and I came to know about so many things I am really thankful to them.

Secondly, we would like to thank our team. During the project, we are all far away from home and have to consider the graduates and employment problems. We encourage and support each other a lot.

Finally, we would like to thanks our parents, who supported us and encouraged us a lot both physically and mentally while studying at the pillai Hoc College of engineering and technology, driving us to study and keep learning.

REFERENCES

- [1] Depression detection from social network data using machine learning techniques, Md. Rafqul Islam, Muhammad AshadKabir , Ashir Ahmed , Abu Raihan M. Kamal , Hua Wang and AnwaarUlhaq, 2018.
- [2] Deep Learning for Depression Detection of Twitter Users, Ahmed HusseinOrabi, PrasadithBuddhitha, Mahmoud HusseinOrabi, Diana Inkpen, 2018.
- [3] Facebook Social Media for Depression Detection in the Thai Community, KantineeKatchapakirin, KonlakornWongpatikaseree, PanidaYomaboot, YongyosKaewpitakkun, 2018.
- [4] Lane DS, et al. From online disagreement to offline action: how diverse motivations for using social media can increase political information sharing and catalyze offline political participation. Soc Media Soc. 2017. <https://doi.org/10.1177/2056305117716274>.
- [5] Yin J, et al. Using social media to enhance emergency situation awareness. IEEE Intell Syst. 2012;27(6):52-9.
- [6] McDougall MA, et al. The effect of social networking sites on the relationship between perceived social support and depression. Psychiatry Res. 2016; 246:223-9.
- [7] Weedon J, Nuland W, Stamos A. Information operations and Facebook version 1.0. 2017. p. 27.
- [8] Krebs F, et al. Social emotion mining techniques for Facebook posts reaction prediction. arXiv Preprint. arXiv:1712.03249. 2017.
- [9] Laganas C, McLeod K, Lowe E. Political posts on Facebook: an examination of voting, perceived intelligence, and motivations. Pepperdine J Commun Res. 2017;5(1):18. 41. Wang Y, Pal A. Detecting emotions in social media: a constrained optimization approach. In: IJCAI. 2015.

- [10] E, Karahalios K. Predicting tie strength with social media. In: Proceedings of the SIGCHI conference on human factors in computing systems. New York: ACM; 2009.
- [11] Hosseinifard B, Moradi MH, Rostami R. Classifying depression patients and normal subjects using machine learning techniques. In: 2011 19th Iranian Conference on Electrical Engineering (ICEE). 2011.
- [12] Islam MR, Kamal ARM, Sultana N, Moni MA, Islam R. Depression detection using K-Nearest Neighbors (KNN) classification technique. In: International conference on computer, communication, chemical, materials and electronic engineering, February 8–9, 2018, Rajshahi, Bangladesh.
- [13] Charoula Angeli, and Sarah K. Howard, "Data mining in educational technology classroom research: Can it make a contribution?," In Computers & Education, Vol. 113, pp.226-242, 2017.
- [14] Wang, Y., Callan, J., & Zheng, B, "Should We Use the Sample? Analyzing Datasets Sampled from Twitter's Stream API," ACM Transactions on the Web (TWEB), Vol 9(3), pp.1-23, 2015
- [15] Sentiment analysis, In Wikipedia. Retrieved November 2, 2017, from https://en.wikipedia.org/wiki/Sentiment_analysis.
- [16] Oscar Araque, Ignacio Corcuera-Platas, J. Fernando Sánchez-Rada, Carlos A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications," In Expert Systems with Applications, Vol. 77, 2017, pp.236-246 .
- [17] S. Loria, Textblob: simplified text processing Secondary Text-Blob: Simplified Text Processing (2014), <https://textblob.readthedocs.io/en/dev/index.html>

IJSER