Sentiment Analysis and Community Detection in Twitter

By – Sai Sandeep – 100769087 and Niranjan Kumar – 100767124

Introduction:

Twitter is the 3rd most used social media platform with more than 340 million active users and generates an average of 500 million tweets a day on trending topics. This huge data generated, and the twitter API connectivity provides us a perfect opportunity for twitter data mining and performs various other activities like sentiment analysis and community detection.

Project Overview:

The objective of this project to implement sentiment Analysis on a global issue like **Global warming, Wildfires, Unemployment** to check if the feeling is positive or negative or neutral. Along with this, we are also implementing community detection for a single twitter user.

Proposed Approach:

For sentiment analysis, we are going to get the tweets using python **tweepy** library that is with a hashtag (#). Then we perform data cleaning process by using **nltk** library we are going to remove the stop words, remove the mentions (@), by using **re** library we can remove the URL if mentioned the tweets and we have to convert the all the tweets to lowercase along with splitting each an every word in a tweet. Later we can find the most common words in all the tweets in a bar graph. By using python **textblob** library we can get the polarity (which ranges from -1 to 1) of the tweet by which we can plot a histogram of all the tweets based on polarity.

For Community detection like the above case, we get the tweets by using the python tweepy library. We are only considering 1-step Neighbourhood that is we consider only 2 levels of friends with a condition that we will not consider friends of friends who are not friends with the main user. We are creating a directed graph with edge weight and node weight. After this, we can create a network of Twitter users. Then we can detect communities based on information flow called info-map or by walk trap.

Datasets:

We are using a custom application to connect to twitter API. Then by using the appropriate hashtags or user ID, we can get all the tweets and user-related information. The only issues with the data are that we must clean the unstructured data before we perform any visualization.

Visualization methods:

We are trying to implement the following visualizations techniques for the proposed project: Word cloud, histograms, bar graphs, social networks graph, and community detection.

References:

[1] C. R. Nirmala, G. M. Roopa and K. R. N. Kumar, "Twitter data analysis for unemployment crisis," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Davangere, 2015, pp. 420-423.

[2] F. Nausheen and S. H. Begum, "Sentiment analysis to predict election results using Python," 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, 2018, pp. 1259-1262.

[3] V. Prakruthi, D. Sindhu and D. S. Anupama Kumar, "Real Time Sentiment Analysis Of Twitter Posts," 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), Bengaluru, India, 2018, pp. 29-34.

[4] Z. Jianqiang and G. Xiaolin, "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis," in IEEE Access, vol. 5, pp. 2870-2879, 2017.

[5] M. Bouazizi and T. Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter," in IEEE Access, vol. 5, pp. 20617-20639, 2017.

[6] A. J. Lam and C. Cheng, "Utilizing Tweet Content for the Detection of Sentiment-Based Interaction Communities on Twitter," 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 2018, pp. 682-691.

[7] V. Arnaboldi, M. Conti, A. Passarella and F. Pezzoni, "Ego networks in Twitter: An experimental analysis," 2013 Proceedings IEEE INFOCOM, Turin, 2013, pp. 3459-3464.

[8] A. Bharadwaju, Y. P. Kumar, K. Anudeep, A. V. Krishna, B. R. Prasad and S. Agarwal, "Real time mining of ego networks for exploring social associations," 2017 Conference on Information and Communication Technology (CICT), Gwalior, 2017, pp. 1-6.

[9] Wendel Silva, Ádamo Santana, Fábio Lobato, and Márcia Pinheiro. 2017. A methodology for community detection in Twitter. In Proceedings of the International Conference on Web Intelligence (WI '17). Association for Computing Machinery, New York, NY, USA, 1006–1009.

[10] M. van Meeteren, A. Poorthuis and E. Dugundji, "Mapping communities in large virtual social networks: Using Twitter data to find the Indie Mac community," 2010 IEEE International Workshop on: Business Applications of Social Network Analysis (BASNA), Bangalore, 2010, pp. 1-8.