

Visualization of Twitter Sentiment during the Period of US Banned Huawei

Nann Hwan Khun, Hninn Aye Thant

University of Technology (Yatanarpon Cyber City)

nannhwankhun@gmail.com, hninyayethant@utycc.edu.mm

Abstract

The polarity analysis of sentiments based on users' expressions on several events has been in much interest for the research. Recently, social media has been popular and it is widely used as a proxy platform to gauge public opinions in real-time. With the growth of microblog sites on the Web, people have started using blog sites like Twitter and other similar social services to express their opinions and emotions on a wide variety of topics. We proposed a visual sentiment analysis framework for US-China trade war related with US banned on Chinese telecoms giant Huawei Technologies. The proposed framework consists of two components, sentiment analysis modeling and geographic visualization. We focus on using Twitter, the most popular microblogging platform, for the task of sentiment analysis by applying lexicon-based approach. This geographic visualization system can help people for better understanding the changes of public sentiment reactions along with the duration and mostly interested regions of Twitter users on this case. In our research, we worked with English language; however, the proposed technique can also be used with any other language.

Key Words- Sentiment Analysis, Social Media, Visualization, United States, China, Huawei

1. Introduction

Sentiment analysis is the process of identifying and analyzing the opinions of users on some specific topics or the texts in attention. Nowadays, there are many popular microblogging sites such as Facebook, Twitter, Instagram, LinkedIn and so on where people can express their emotions and opinions and discuss about every topic that we can imagine. It has been useful in several domains like economic, education, health and even in political domain. Public opinion expressions are the main focus of a certain amount of research on social media. In our research, the social media service that we focus on is Twitter that lets users to send and read text-based messages which are usually called 'tweets'. There are different kinds of evidence exist that Twitter can be used to identify the

sentiment or emotion of tweets based on text classification [1].

According to the strong usage and development of social media, tracking public opinions becomes easier, faster and more reliable than using the traditional public surveys or polls. Moreover, social media may even give useful information whereas traditional surveys are shortened because of limited financial amounts. Information from social media can be combined with traditional survey techniques as it gives solid measurements of behavior over time, while taking advantage of huge population sample sizes. The former US President, Obama, administration used sentiment analysis to investigate public opinion related with policy announcements and campaign messages ahead of 2012 presidential election and found that it was very useful. So, the applications of sentiment analysis are vast and powerful.

Since microblogging can be used for sentiment analysis purposes, we use microblogging and Twitter as a specific corpus for sentiment analysis and opinion mining in our system for the following reasons:

- People usually express their opinions and emotions for different cases on the several microblogging platforms and so, it becomes a valuable source of people's opinions and feelings.
- Twitter's corpus is growing every day and can be arbitrarily large as Twitter contains voluminous number of text posts.
- People from different backgrounds such as regular users to celebrities, company CEOs, politicians and even country presidents usually become Twitter's audiences. Therefore, it is possible to collect text posts of users from different interest and age groups on current affairs.
- It is also possible to collect posts in different languages as Twitter is used in worldwide and the users are from many different countries.

The new user-centralized Web services hold a vast amount of data created by various users. Users are now considered as co-creators of web content and the social media is now becoming a major part of the Web. The user contributions to social media range from blog posts, tweets, reviews and photo/video uploads etc. Most of the data on the Web is unstructured text. Opinions uploaded

in social media as reviews or posts make an important and interesting research area. The new challenging task for the development in accessibility of opinion resource such as product reviews, blog reviews, and social network posts is to dig into huge volume of texts and design suitable algorithms to understand the opinion of users.

The US-China trade war is recently evolving into a fight over which of the world's two largest economies will control the future of technology and telecommunications. The Trump administration has stepped up its attacks on Huawei. On May 15, 2019, President Trump issued an executive order banning all U.S. companies from utilizing information and communications technology from any party considered a national security threat. The order also declared a national emergency related to this matter. The U.S. Commerce Department also added Huawei and 70 of its affiliates to its existing "Entity List" and this blacklist bars anyone on it from purchasing parts and components from U.S. companies unless they have prior government approval.

US President Donald Trump's ban on Chinese telecoms giant Huawei Technologies, and its implications for 5G development, poses a more serious threat to the global economy than higher tariffs, according to economists. The US-China tech competition will be more intense and drawn out, significantly outlasting the trade tensions.

We proposed a geographic visualization for sentiment analysis to reveal public opinion toward US-China trade war expressed on Twitter over one month period. Originating locations of the Twitter posts are visualized on the Geo-map and each country and their related tweet counts on this case are also extracted.

2. Related Work

In general, there are three types of approaches for texts sentiment analysis: (a) using a machine learning based classifier such as Naïve Bayes, SVM or Neural Networks with suitable feature selection scheme; (b) using the obtained and labeled relevant n-grams of the text and the document with unsupervised semantic orientated scheme; and (c) using the publicly available lexicon dictionaries that provides positive, negative and neutral scores for each word. Some of the relevant past works on sentiment classification can be found in [2] and [3].

Dictionaries associated with sentiment polarity for words are called opinion lexicons. Their use in opinion mining research derives from the hypothesis according to the consideration of each individual word as a unit of opinion information and thus can provide suggestions to review sentiment and subjectivity. In [4], the authors counted positive and negative terms found in a review by

applying SentiWordNet lexicon and determine sentiment based on which class receives the highest score.

Sentiment analysis and emotion analysis are the tasks which identify the attitude and emotion classes of the investigated document [5]. The analysis is determined on the several different domains like educational, political, environmental and social etc. Feature-based classification approach of sentiment analysis has been used from the first application of sentiment analysis of movie reviews to current emotion analysis applications on social media [6]. Besides, visual analytics is widely used in social media data analysis such as information diffusion, business analysis and geographic analysis.

The proposed research is all about sentiment analysis using lexicon-based approach which determines the sentiment or polarity of opinion by looking at the opinion words in each document or sentence.

One of the simple approaches to perform sentiment analysis is depended on the function of opinion words in context where opinion words are words that are generally used to express positive or negative sentiments, e.g., "good" and "bad". These approaches basically use an opinion lexicon to describe and diagnose sentiment polarity (positive, negative or neutral). Such kind of approach is called the lexicon-based approach [7].

In this paper, we investigate the feasibility of accessing public risk perception using social media analysis which has recently been used to estimate trends of interest such as stock prices, movie sales, political mood and influenza rates, by applying a similar methodology to assess the public's level of concern toward the trade tension.

The attitudes toward US-China trade war expressed on Twitter are examined as a case study. Over 110,000 Twitter messages referencing this trade war between May 15th and June 15th, 2019 are collected and are split automatically between three sets of texts:

- Texts containing positive opinions.
- Texts containing negative opinions.
- Objective texts that only state a fact or do not express any emotions or opinions.

We specifically look at how message volume changes and how sentiment varies over time and location. A demographic analysis of the proportion of concerned messages shows significant differences by visualizing on geo-map.

The remainder of the paper is organized as follows: First, we describe the proposed method in details. Next, we present sentiment classification experiments using publicly available lexicons (dictionaries). Finally, we analyze demographic variation in levels of concern and visualize on the real geo-map.

3. Proposed Method

The proposed system is mainly composed of four components: data collection, data preprocessing, sentiment classification and visualization. The architectural overview of the proposed system is shown in Figure 1.

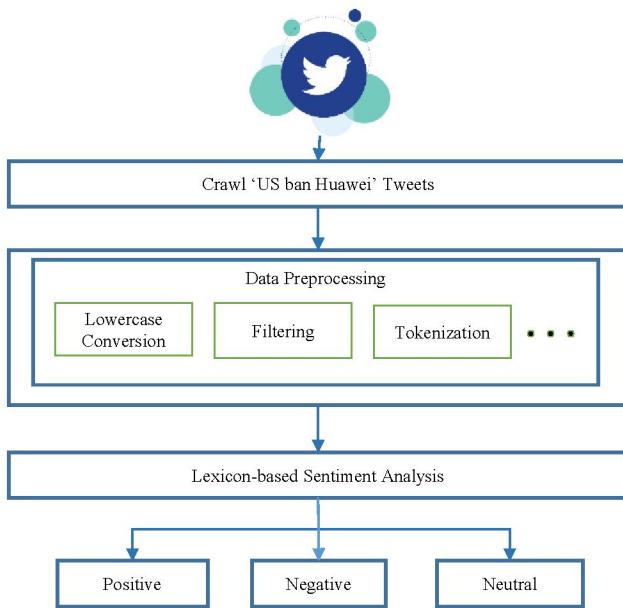


Figure 1. Architectural Overview of the Proposed System

3.1. Data Collection

The amount of data that one has to deal has exploded to unimaginable levels in the previous decade. Since big data analysis mainly involves collecting data from diverse sources, it becomes available to be used up by analysts and can finally deliver useful data to the organizational business. So, it can be said that the core of big data analysis is the process of transforming large amounts of unstructured raw data from different sources to a useful data as a product for organizations. The proposed system is working on collecting data from Twitter and structuring it for sentiment analysis model.

The Twitter data programmatically can be accessed from the Twitter API that interacts with the created application inside Twitter. OAuth [8] protocol is used for Authentication in order to search for particular Tweets. Then, we get a consumer key, consumer secret, access token and access token secret that should always be kept as private. After obtaining these four keys, we connect to the Twitter Streaming API [9] and start downloading Tweets. One tweet contains at most 140 characters and these tweets are from various Twitter users' status. In this

way, the system crawls the real tweets related with US-China trade tensions on Huawei case.

3.2. Data Preprocessing

For doing analytical process, data preprocessing is very important stage, which transforms the raw data into the valuable data. Real-world data has so much noise and is often incomplete, inconsistent, and lacking in certain behaviors or trends. Every social media analysis is required to pass this data preprocessing stage in order to get the desired goal. In this system, six preprocessing steps are performed using Natural Language Toolkit techniques.

- Conversion of all text to lowercase - this helps to minimize multiple copies of the exact similar word.
- Removal of usernames, links and special symbols – because they do not contain important information in text data analysis.
- Word-tokenization - this divides the text into a sequence of words and it plays a large part in the process of text data analysis.
- Stop-words Removal – removal of uninformative and article words like a, an, the etc. This is mainly to improve the system accuracy and to reduce the required computation time.
- Lemmatization – this converts the word into its original word, that we call lemma.
- Spelling Correction – this step is also performed since tweets usually contain misspelt words.

3.3. Sentiment Classification

Sentiment Analysis, or Opinion Mining, is a sub-field of Natural Language Processing (NLP). The general process of sentiment analysis is identifying and extracting opinions within a given text. The major goal of sentiment analysis is to gauge the sentiments, evaluations, attitudes and emotions of a speaker/ writer based on the computational treatment of subjectivity in a text.

The lexicon-based approach depends on opinion (or sentiment) dictionary with each word that express positive or negative sentiment. Researchers have compiled sets of opinion words and phrases for adjectives, adverbs, verbs and nouns respectively. To create (much less, validate) a complete sentiment lexicon is a labor intensive and sometimes error prone process, so it is no doubtful that many opinion mining researchers and practitioners rely so massively on existing lexicons as primary resources. Similarly, we used standardized English opinion lexicon that is an extensible English language lexicon for POS tagging with around 110K words and performed lexicon-based sentiment analysis. The system design of our proposed system is shown in Figure 2.

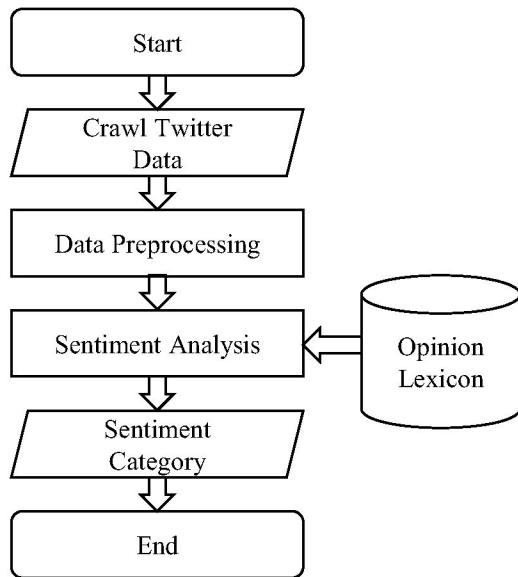


Figure 2. System Design

3.4. Visualization

Data visualization generally aims to help people understand the implication of data by placing it in a visual situation. The abstract information for data analysis and communication is displayed in a graph which leads to understanding for action. In this system, we show not only how the sentiment of people accompanying with this tension on real geographic visual but also the locations of Twitter users on the geo-map.

A novelty graphical display or visual representation of text data or to visualize free form text is shown as Word Cloud. Figure 3 is a Word Cloud representation in order to have a global vision related with US banned Huawei Twitter dataset. The more the word occurs, the bigger the font size is.

Twitter supports several types of APIs, and tweets generally come as JSON objects together with the tweet text along with metadata, such as time, location with coordinates which are associated with the tweet (if provided by the user). Since there are four primary ways in which geolocation is commonly performed on Twitter users and messages, we used JSON ‘Place’ object that encodes a location associated with the tweet. Figure 4 is obtained based on these geographic coordinates. This system shows the real geolocation of Twitter users, who posted about this Huawei case, with red spots on the geo-map as shown in Figure 4. However, it does not cover all the users since only a few tweets contain geolocations.

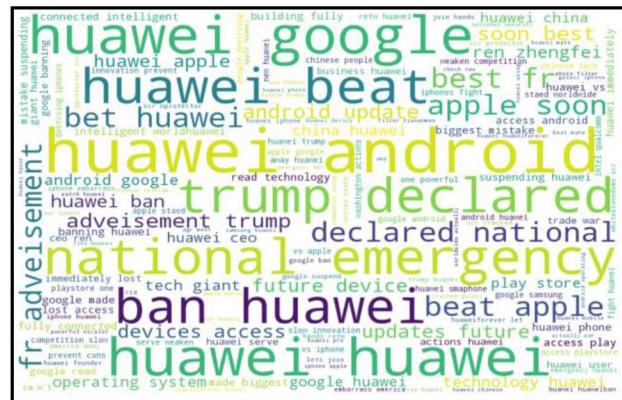


Figure 3. Word Cloud Visualization

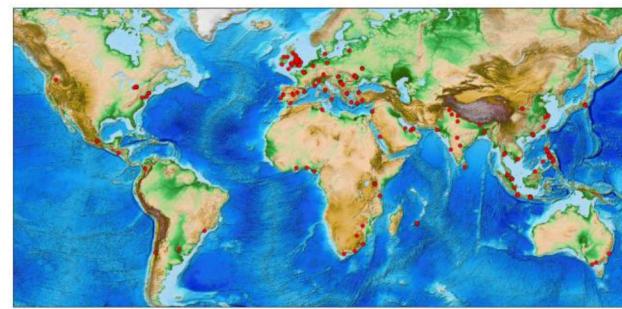


Figure 4. Visualization on Geo-map during US banned Huawei

4. Experimental Results and Discussion

Our system works on the ‘US banned Huawei’ Twitter corpus that consists of 115775 tweets during one month (15th May to 15th June, 2019). This whole dataset is used for geographic visualization related with this case but our experiment for sentiment analysis works on 74627 tweets. Our system defines sentiment score for each of the tweets. The score is calculated according to the number of positive (+1) and negative words (-1) in a tweet, as well as the number of negation words (such as no, not; -1). The system shows the sentiment analysis result of this trade tension by using bar graph as shown in Figure 5.

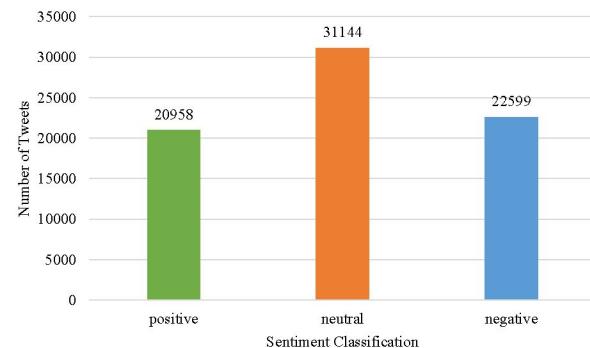


Figure 5. Sentiment Analysis Result of “US banned Huawei” Dataset

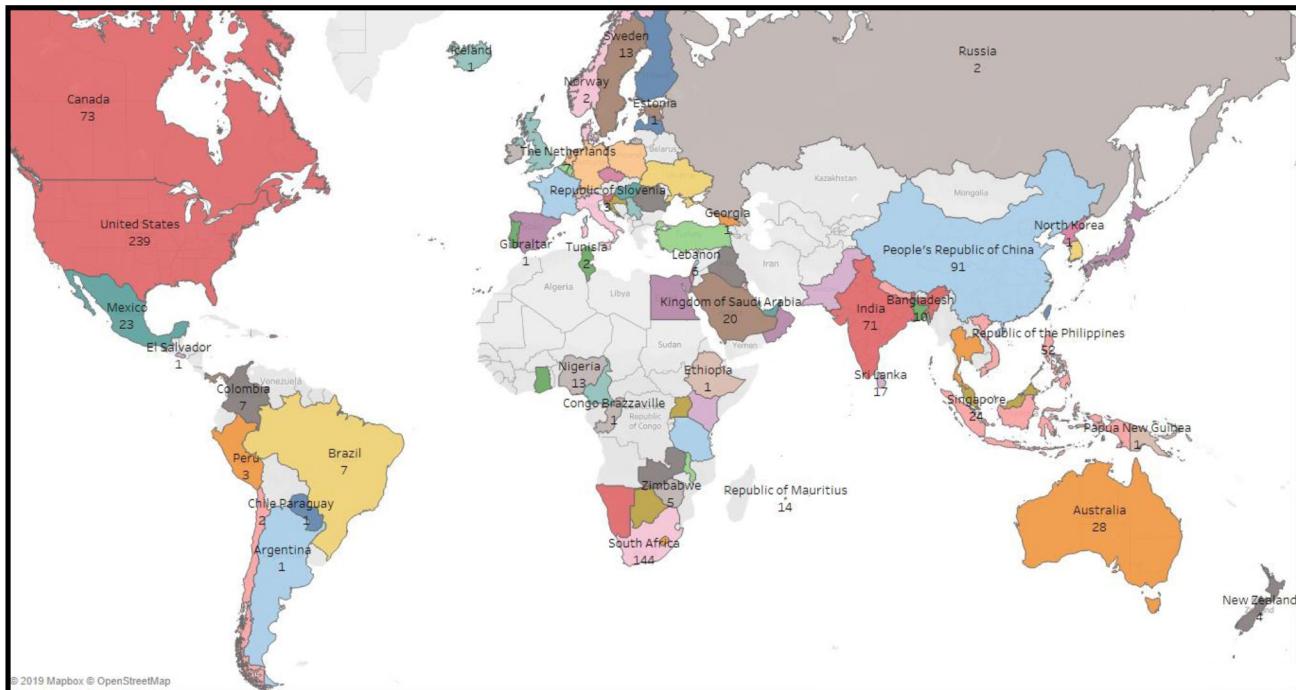


Figure 6. Interested Countries and their Tweet Counts on “US banned Huawei” Dataset using ‘Country’ Object

From this result, we can see that people who agree on this “US banned Huawei” case is less than both do not agree persons and people who have no opinions on this case.

Our geographic system also extracted which countries are mostly interested on this case using Twitter ‘Country’ object as shown in Figure 6 by using Tableau

visualization tool. It is obvious that the number of tweet counts is much less since most of the tweets leave blank in ‘Country’ object. Therefore, we tried another way to know the original location of each tweet by manually analyzing ‘location’ object of each tweet. The geographic visualization result of each country and their tweet counts by using ‘location’ object is shown in Figure 7.

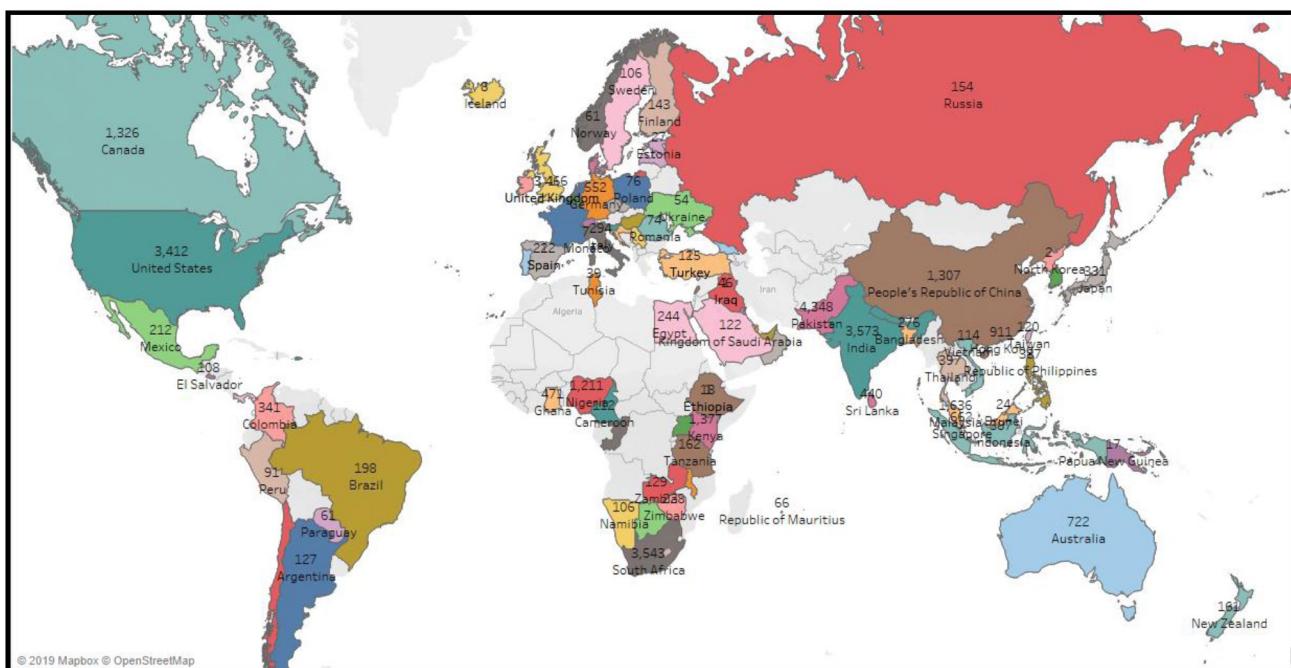


Figure 7. Interested Countries and their Tweet Counts based on Tweets’ Locations

There is no doubt that most of the tweets were came from both United States and China since this Huawei case is one of the trade tensions between US and China. We also noticed another interesting thing that is voluminous amounts of tweets were also came from Canada, United Kingdom, India, Pakistan and South Africa. Their tweet amount is even more than that from China according to ‘location’ object. This may be because of few Twitter users from China.

The original location (based on ‘coordinate’ object) of each sentiment analysis result is plotted on the geographic visualization as shown in Figure 8 with different color cycles representing green for positive, blue for negative and orange for neutral respectively. We can obviously see that the most scattered color cycle over the whole geo-map is blue cycle that represents for negative emotion and so it can be said that most of the people seen from negative point of view on this case generally, it is true by accompanying with other trade tension cases.

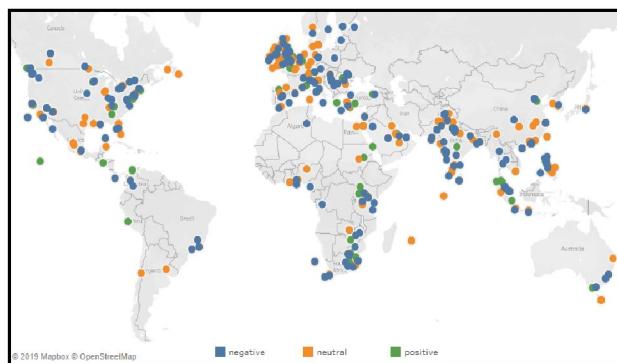


Figure 8. Geographic Visualization of Sentiment Analysis Result

5. Conclusion

In this paper, we presented lexicon-based Twitter sentiment analysis system with geographic vision during the ‘US ban Huawei’ period. This analysis result can be useful not only for the politicians but also for the economists to reveal public opinion. Our system is only focused on one month during the ‘US ban Huawei’ period. Later, this system can also be broaden for time-series sentiment analysis along with this trade war tension days.

There are also some kinds of work for further extending on this system by combining with supervised machine-learning methods to become hybrid method and using big data solutions to get results that are more accurate. Moreover, by using various lexicon resources, we can also aware which lexicon is most suitable on such this case. By incorporating with language professionals, our system can also be extended for different languages.

6. References

- [1] K. Glass and R. Colbaugh, “Estimating the sentiment of social media content for security informatics applications”, in IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 65-70, Beijing, 2011.
- [2] K. T. Durant and M. D. Smith, “Mining sentiment classification from political web logs”, In Proc. of WEBKDD’06, ACM, August, 2006.
- [3] P.Walia, Marisha, V.K.Singh, and M.K,Singh, “Evaluating machine learning and unsupervised semantic orientation approaches for sentiment analysis of textual reviews”, In Proc. of Intl. Conf. on Computational Intelligence and Computing Research, Coimbatore, India, 2012.
- [4] B. Ohana and B. Tierney, “Sentiment Classification of Reviews Using SentiWordNet”, 9th IT & T Conference, Dublin Institute of Technology, Dublin, Ireland, 22nd.-23rd. October, 2009.
- [5] C. Strapparava and R. Mihalcea, “Learning to identify emotions in text”, In Proc. of the 2008 ACM Symposium on Applied Computing, pp. 1556-1560, New York, 2008.
- [6] B. Pang and L. Lee, “Opinion mining and sentiment analysis”, Foundations and Trends in Information Retrieval, vol.2, no. 1-2, pp. 1-135, 2008.
- [7] A. Aue and M. Gamon, “Customizing Sentiment Classifiers to New Domains: a Case Study”, RANLP 2005.
- [8] <https://developer.twitter.com/en/docs/authentication/overview/oauth> [Accessed: 2019-04-23]
- [9] <https://apifriends.com/api-streaming/understand-twitter-streaming-api/> [Accessed: 2019-5-14]