Twitter Sentiment Analysis

Rishi Ittan[1], Shaurya Chawla[2]

[1,2]*Student, Department of Information Technology, Maharaja Agrasen Institute of Technology, New Delhi, India*

Abstract: **Social networks are the principal resources to gather data about people's opinion and sentiments in the direction of extraordinary subjects as they spend hours day by day on social medias and share their opinion. In this technical paper, we exhibit the utility of sentimental analysis and how to connect to Twitter and run sentimental analysis queries. We run experiments on unique queries from politics to humanity and exhibit the interesting results. We realized that the neutral sentiment for tweets are extensively high which virtually suggests the boundaries of the contemporary works.**

Keywords: **Twitter sentiment analysis, Social Network analysis.**

1. Introduction

Opinion and sentiment mining is a necessary research place due to the fact due to the huge number of each day posts on social networks, extracting people's opinion is a difficult task. About 90 percent of today's statistics has been supplied at some point of the final two years and getting insight into this large scale data is no longer trivial [17], [18].

Sentimental evaluation has many purposes for extraordinary domains for example in organizations to get remarks for merchandise by means of which agencies can analyze user's comments and reviews on social medias.

Opinion and sentimental mining has been well studied in this reference and all distinct strategies and lookup fields have been mentioned [10]. There are additionally some works have been performed on Facebook [19]-[23] sentimental analysis on the other hand in this paper we in most cases focal point on the Twitter sentiment analysis.

For a larger texts one answer may want to be recognize the text, summarize it and supply weight to it whether or not it is positive, negative or neutral. Two fundamental strategies to extract textual content summarization are an extractive and abstractive method. In the extractive method, phrases are extracted from the original textual content to generate a summary. In an abstractive method, tries to analyze an internal language illustration and then generates summary that is greater similar to the summary performed by human.

Text understanding is a massive hassle to solve. Some machine learning techniques, which include a variety of supervised and unsupervised algorithms, are being utilized. There are distinct tactics to generate summary. One strategy should be rank the importance of sentences within the text and then generate summary for the textual content primarily based on the importance numbers. There is some other strategy called end-to-end generative models. In some domain like image

recognition, speech recognition, language translation, and question-answering, the end-to-end technique performs better.

Some works have used an ontology to recognize the text [1]. In the phrase level, sentimental analysis system ought to be in a position to recognize the polarity of the phrase which is mentioned through Wilson, et.al [9]. Tree kernel and feature based model have been utilized for sentimental evaluation in twitter through Agarwal and et.al [11]. SemEval-2017 [12] additionally shows the seven years of sentimental analysis in twitter tasks. Since tweets in Twitter is a particular text not like a normal textual content there are some works that address this problem like the work for quick informal texts [13]. Sentimental evaluation has many applications in information [14].

In this paper, we will discuss social network analysis and the significance of it, then we discuss Twitter as a wealthy aid for sentimental analysis. In the following sections, we exhibit the high-level summary of our implementation. We will show some queries on special topics and exhibit the polarity of tweets.

2. Social network analysis

Social network analysis is the study of people's interactions and communications on exclusive matters and in modern times it has obtained greater attention. Millions of people provide their opinion of distinct matters on an everyday basis on social medias like Facebook and Twitter. It has many purposes in exclusive areas of research from social science to commercial enterprise [3].

Twitter in modern times is one of the popular social media which in accordance to the statistics [4] currently has over 300 millions accounts. Twitter is the wealthy supply to analyze about people's opinion and sentimental evaluation [2]. For each tweet it is vital to determine the sentiment of the tweet whether is it positive, negative, or neutral.

Another challenge with twitter is only 140 characters is the drawback of each tweet which reason human beings to use phrases and works which are not in language processing. Recently twitter has extended the textual content barriers to 280 characters per each tweet.

3. Twitter sentiment analysis

Social networks are wealthy platforms to analyze about people's opinion and sentiment concerning unique topics as they communicate and share their opinion actively on social medias such as Facebook and Twitter. There are different

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792

247

opinion-oriented records gathering systems which intention to extract people's opinion concerning distinct topics. The sentiment-aware systems these days have many applications from commercial enterprise to social sciences.

Since social networks, particularly Twitter, consists of small texts and people might also use different words and abbreviations which are challenging to extract their sentiment through contemporary Natural Language processing systems easily, consequently some researchers have used deep learning and machine learning strategies to extract and mine the polarity of the text [15]. Some of the popular abbreviations are FB for Facebook, B4 for before, OMG for oh my god and so on. Consequently, sentimental analysis for brief texts like Twitter's posts is difficult [8].

4. Design and implementation

This technical paper reviews the implementation of the Twitter sentiment analysis, through utilizing the APIs provided by Twitter itself.

There are incredible works and equipment focusing on text mining on social networks. In this research the wealth of available libraries has been used.

The approach to extract sentiment from tweets is as follows:
- Start with downloading and caching the sentiment dictionary.
- Download twitter testing data sets, input it into the program.
- Remove the stop words to clean the tweets.
- Tokenize and feed into the program every word in the dataset.

Compare it to positive sentiments and negative sentiments in the dictionary for each word. Then positive count increment or negative count increment.

Based on the count either positive or negative, we can finally analyze percentage of sentiment to decide the polarity.

Researchers have done distinctive sentimental analyses on Twitter for diverse purposes for example the work outlined by Wang, et.al [5] is a real-time twitter sentiment analysis of the presidential race.

Fig. 1 shows the sentimental analysis algorithm at the high level. As it can be seen in the algorithm, we have different procedures to connect the twitter API, fetch the tweets, tweet cleaning or remove stop words, classify tweets which means get the polarity of the tweet, and finally return the results.

A. Implementation

In this paper, we used python to implement sentiment analysis. Some packages have utilized including tweepy and textblob. We can install the required libraries by following commands:
- pip install tweepy
- pip install textblob

The second step is downloading the dictionary by running the following command:

python -m textblob.download_corpora.

The textblob is a python library for text processing and it uses NLTK for natural language processing [6]. Corpora is a large and structured set of texts which we need for analyzing tweets.

B. Connect to Twitter using APIs

To connect to Twitter and query latest tweets, we need to create an account on twitter and define an application. Users need to go to the apps.twitter.com/app/new and generate the api keys. The Application settings is shown in the fig. 2. Due to the security reasons the api keys are not shown.
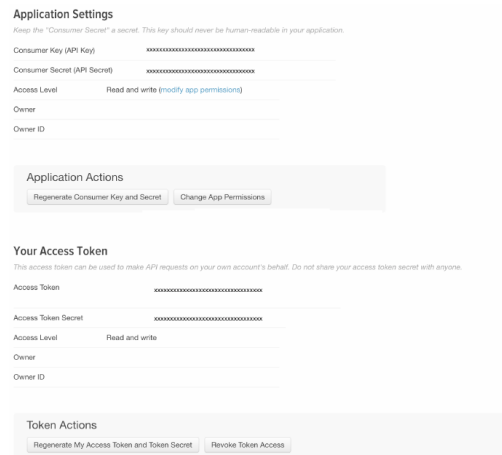

Fig. 2. Twitter application management

C. Sample results

Following shows the sample output of the program for the 'fake news' as a query based on the last 300 tweets from Twitter.

Positive tweets percentage: 16.39 %
Negative tweets percentage: 72.13 %


```
Algorithm 1 Extract Twitter sentiment
 1: procedure TWITTER-CONNECTION()
 2:     consumer − key =′ xxxxxxxx′
 3:     consumer − secret =′ xxxxxxxx′
 4:     access − token =′ xxxxxxxx′
 5:     access − token − secret =′ xxxxxxxxx′
 6:     self.auth = OAuthHandler(consumer − key, consumer − secret)
 7:     self.auth.set − access − token(access − token, access − token − secret)
 8:     self.api = tweepy.API(self.auth)
 9: end procedure
10:
11: procedure TWEET-CLEANING(t)
12:     tweet = t.remove − Stop − words
13:     Return tweet
14: end procedure
15:
16: procedure TWEET-CLASSIFICATION(t)
17:     t = Tweet − Cleaning(t)
18:     tweet − polarity = t.sentiment.polarity
19:     tweet − polarity
20: end procedure
21:
22: procedure GET-TWEETS(q, count)
23:     fetched − tweets = self.api.search(q = query, count = count)
24:     Return fetched − tweets
25: end procedure
26:
27: procedure MAIN()
28:     st = SentimentalTwitter()
29:     tweets = st.fetch − tweets(query =′ politics′, count = 300)
30:     PositiveTweets = tweetsthatsentiment =′ positive′
31:     NegativeTweets = tweetsthatsentiment =′ negative′
32:
33:     for tweet t in PositiveTweets do
34:         print(t)
35:     end for
36:     for tweet t in NegativeTweets do
37:         print(t)
38:     end for
39: end procedure
```

Fig. 1. Extract sentiment from Twitter

248

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792

Neutral tweets percentage: 11.47 %

Positive tweets:

tweet: @Nigel_Farage @PoppyLegion Least we forget: Farage is rich. Brexit makes him richer. He is establishment. He is a l… https://t.co/FhZSCBVHJs

tweet: @kirk0071 @Scavino45 @WhiteHouse @POTUS @realDonaldTrump Thanks for the good belly laugh this morning. Your HateTru… https://t.co/AWHXoC84LJ

tweet: @rolandsmartin Roland I like you brother but you really need to distant yourself from Donna Brazile,she's been comp… https://t.co/zqRCsVu98d

Negative tweets:

tweet: RT @Independent: If you saw these tweets, you were targeted by Russian Brexit propaganda https://t.co/Cc8IvQApbY

tweet: Behind Fox News' Baseless Seth Rich Story: The Untold Tale https://t.co/TXcDP1oQ5H

tweet: RT @JackPosobiec: Fake news called the Poland indpendence day parade a "Nazi march." Sick https://t.co/OZA3xUopl1

Table 1 shows the sentimental analysis results based on different queries including movie, politics, fashion, and fake news. The bar chart, as shown in figure 3, illustrates the data based on the results we got form this step. If we run the program in different times we may get different results, small variance, based on the tweets we fetch. We run the program three times and these results are the average of the outputs.

Table 1
Sentiment analysis results

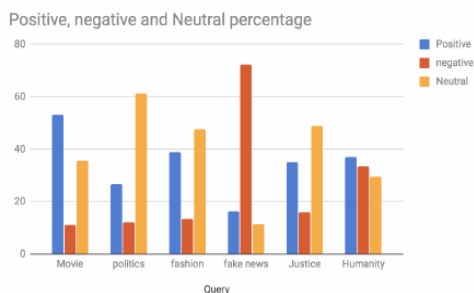| Query	Positive	negative	Neutral
Movie	53	11.1	35.8
politics	26.6	12.2	61.1
fashion	38.8	13.3	47.7
fake news	16.3	72.1	11.4
Justice	35.2	15.9	48.8
Humanity	36.9	33.3	29.7

Fig. 3. Sentiment results on different queries

As it can be clearly seen in the table and diagram the percentage of the neutral tweets are significantly high. This is also important to mention that depends on the data of the experiment we may get different results as people's opinion may change depends on the world circumstances for example fake news as it becomes the world of the year in 2017. For some queries, the neutral tweets are more than 60% which clearly shows the limitation of the current works.

5. Conclusion

In this technical paper, we discussed the importance of social network analysis and its applications in different areas. We focused on Twitter as and have implemented the python program to implement sentiment analysis. We showed the results on different daily topics. We realized that the neutral sentiments are significantly high which shows there is a need to improve Twitter sentiment analysis.

References

[1] Boguslavsky, I. (2017). Semantic Descriptions for a Text Understanding System. In Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2017) (pp. 14-28).

[2] Pak, A., & Paroubek, P. (2010, May). Twitter as a corpus for sentiment analysis and opinion mining. In LREc (Vol. 10, No. 2010).

[3] Scott, J. (2011). Social network analysis: developments, advances, and prospects. Social network analysis and mining, 1(1), 21-26.

[4] Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012, July). A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 System Demonstrations (pp. 115-120). Association for Computational Linguistics.

[5] TextBlob, 2017, https://textblob.readthedocs.io/en/dev/

[6] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1-135.

[7] Dos Santos, C. N., & Gatti, M. (2014, August). Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts

[8] Wilson, T., Wiebe, J., & Hoffmann, P. (2005, October). Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the conference on human language technology and empirical methods in natural language processing (pp. 347-354). Association for Computational Linguistics.

[9] Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

[10] Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011, June). Sentiment analysis of twitter data. In Proceedings of the workshop on languages in social media (pp. 30-38). Association for Computational Linguistics.

[11] Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), (pp. 502-518).

[12] Kiritchenko, S., Zhu, X., & Mohammad, S. M. (2014). Sentiment analysis of short informal texts. Journal of Artificial Intelligence Research, 50, 723-762.

[13] Balahur, A., Steinberger, R., Kabadjov, M., Zavarella, V., Van Der Goot, E., Halkia, M., ... & Belyaeva, J. (2013). Sentiment analysis in the news. arXiv preprint arXiv:1309.6202.

[14] Poria, S., Cambria, E., & Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (pp. 2539-2544).

[15] Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, 527-541.

[16] Bagheri, Hamid, and Abdusalam Abdullah Shaltooki. "Big Data: challenges, opportunities and Cloud based solutions." International Journal of Electrical and Computer Engineering 5.2 (2015): 340.

[17] Bagheri, Hamid, Mohammad Ali Torkamani, and Zhaleh Ghaffari. "Multi-Agent Approach for facing challenges in Ultra-Large Scale

249

International Journal of Research in Engineering, Science and Management
Volume-2, Issue-4, April-2019
www.ijresm.com | ISSN (Online): 2581-5792

systems." International Journal of Electrical and Computer Engineering 4.2 (2014): 151.

[18] Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. Computers in Human Behavior, 31, 527-541.

[19] Feldman, R. (2013). Techniques and applications for sentiment analysis. Communications of the ACM, 56(4), 82-89.

[20] Dasgupta, S. S., Natarajan, S., Kaipa, K. K., Bhattacherjee, S. K., & Viswanathan, A. (2015, October). Sentiment analysis of Facebook data using Hadoop based open source technologies. In Data Science and Advanced Analytics (DSAA), 2015. 36678 2015. IEEE International Conference on (pp. 1-3). IEEE.

[21] Trinh, S., Nguyen, L., Vo, M., & Do, P. (2016). Lexicon-based sentiment analysis of Facebook comments in Vietnamese language. In Recent developments in intelligent information and database systems (pp. 263-276). Springer International Publishing.

[22] Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre- processing in sentiment analysis. Procedia Computer Science, 17, 26-32.

[23] Statista, 2017, https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/