IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

Sentiment Analysis on Twitter

Lokesh Patidar[1*], Raghavendra Nayaka P.[2],Vaibhav Malviya[3], Ashwini[4], Varshitha TR[5]

[1,2,3,4,5] School of Computing and Information Technology, Reva University, Karnataka, India

Corresponding Author: lokeshpatidar63@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7si14.420423 | Available online at: www.ijcseonline.org

***Abstract*—**Twitter is an online miniaturized scale blogging and person to person communication stage which enables clients to compose short notices of most extreme length 140 characters (280 characters for confirmed records). This task tends to the issue of conclusion investigation in twitter; that is ordering tweets as indicated by the notion communicated in them: positive, negative or nonpartisan. It is a quickly growing administration with more than 500 million enlisted clients - out of which 330 million are dynamic clients and half of them sign on twitter once a day - producing almost 500 million tweets for each day. Because of this huge measure of use we would like to accomplish an impression of open assessment by breaking down the conclusions communicated in the tweets. Investigating the open slant is vital for some applications, for example, firms endeavouring to discover the reaction of their items in the market, foreseeing political races and anticipating financial wonders like stock trade.

***Keywords*—**Social Network, Sentiment Analysis, Big Data, Applications

I. INTRODUCTION

Sentiment analysis utilizes information mining procedures and strategies to concentrate and catch information for examination so as to recognize the abstract sentiment of a report or gathering of records, similar to blog entries, audits, news articles and web based life channels like tweets and notices.

We have worked with twitter since we feel it is a superior estimate of open supposition rather than traditional web articles and web online journals. The reason is that the measure of pertinent information is a lot bigger for twitter, when contrasted with customary blogging destinations. In addition, the reaction on twitter is increasingly speedy and furthermore progressively broad (since the quantity of clients tweet's identity generously more than the individuals who compose web writes every day). Slant examination of open is very basic in full scale financial marvels like foreseeing the securities exchange rate of a specific firm. This should be possible dissecting by and large open opinion towards that firm as for time and utilizing financial devices for finding the relationship between open slant and the company's securities exchange esteem. Firms can likewise assess how well their item is reacting in the market, which territories of the market is it having a great reaction and in which a negative reaction (since twitter enables us to download stream of geo-labelled tweets for specific areas). On the off chance that organizations can get this data they can break down the explanations for topographically separated reaction, thus they can advertise their item in a more upgraded way by searching for fitting arrangements like making reasonable market sections. Foreseeing the aftereffects of well-known political races and surveys is likewise a developing application to supposition examination. One such examination was directed by Tumasjan in Germany for anticipating the result of government decisions in which reasoned that twitter is a decent impression of disconnected opinion [1].

II. RELATED WORK

Twitter Sentiment Analysis: The Good the Bad and the OMG! In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011. By Efthymios Kouloumpis, Theresa Wilson and Johanna Moore[2]**.**
The bag of-words display is a standout amongst the most broadly utilized element demonstrate for practically all content order errands because of its straightforwardness combined with great execution. The model speaks to the content to be delegated a pack or accumulation of individual words with no connection or reliance of single word with the other, for example it totally slights language and request of words inside the content. This model is likewise well known in supposition examination and has been utilized by different specialists. The absolute soonest work in this field ordered content just as positive or negative, expecting that every one of the information gave is abstract. While this is a decent supposition for something like film audits yet while breaking down tweets and sites there is a great deal of target content

we need to consider, so joining nonpartisan class into the order procedure is presently turning into a standard.

Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In the Annual Meeting of Association of Computational Linguistics: Human Language Technologies (ACL-HLT), 2005. By Theresa Wilson, Janyce Wiebe and Paul Hoffmann [3].

Sentiment-analysis of in the space of miniaturized scale blogging is a moderately new research theme so there is still a ton of space for further research here. Average measure of related earlier work has been done on Sentiment-analysis of client audits, archives, web journals/articles and general expression level sentiment analysis. These contrast from Twitter basically due to the furthest reaches of 140 characters for each tweet which powers the client to express feeling compacted in exceptionally short content. The best outcomes came to in sentiment order utilize directed learning procedures, for example, Naive Bayes and Support Vector Machines, however the manual marking required for the regulated methodology is pricey.

Other than from these much work has been done in investigating a class of highlights appropriate just to small scale blogging area. Nearness of URL and number of capitalized words/alphabets in a tweet have been investigated by Koulompis et al. furthermore, Barbosa et al., Koulompis also reports positive outcomes for utilizing emojis and internet slang words as highlights.

The most normally utilized characterization strategies are the Naïve Bayes Classifier and State Vector Machines. Naïve Bayes strategies are a lot of managed learning calculations dependent on applying Bayes' hypothesis calculation with the 'naïve' presumption of autonomy between each pair of highlights.

III. METHODOLOGY

The way toward structuring our sentiment analyzer, we partitioned our work into five fundamental classes. They are as follows:

1. Authentication
2. Data Collection
3. Data Pre-processing
4. Analysis

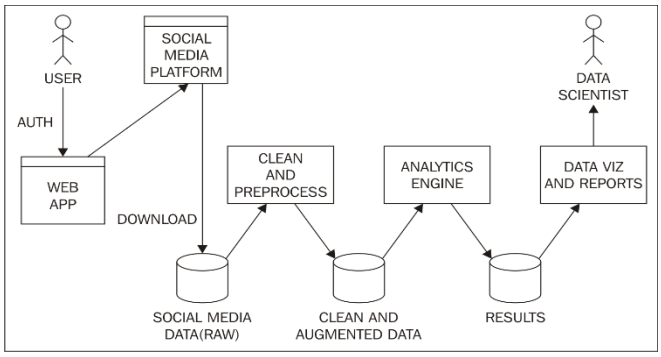

Figure 1the overall process of Twitter mining

Authentication

To cooperate with the Twitter APIs, we need a Python client that executes the distinctive calls to the API itself. There are a few choices as should be obvious from the official documentation, none of them are formally kept up by Twitter and they are upheld by the open source community. The initial segment of the communication with the Twitter API includes setting up the confirmation.

To setup authentication we created a web app on https://apps.twitter.com/and got our keys and access tokens using our twitter credentials. Now we needed to create a program (twitterClient.py) which can authenticate and retrieve data from twitter API.

For this process there are various python libraries like Tweepy, Twython etc., we have used Tweepy which provides easy access to the Twitter API.

Data Collection

Twitter provides more than a single API,we can categorize our options into two classes: REST APIs and Streaming API.

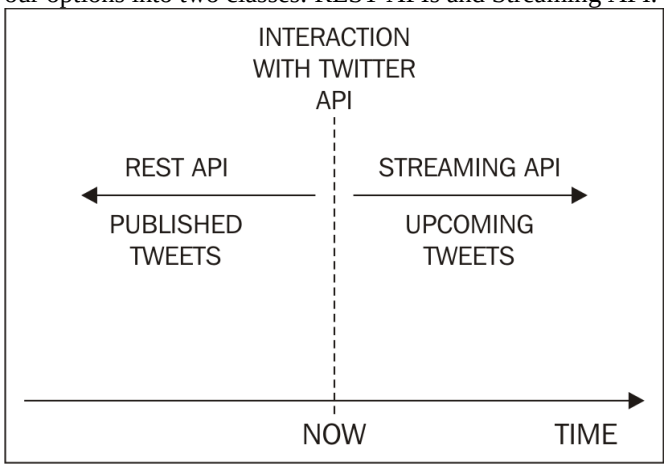

Figure 2 Two ways of collecting data from Twitter

All the REST APIs only allow you to go back in time. Then again, the Streaming API investigates what's to come.

When we open a connection, we can keep it open and go ahead in time. By keeping the HTTP connection open, we

can recover every one of the tweets that match our filter criteria, as they are published.
Data from Twitter is collected in JSON (JavaScript Object Notation) format which is a syntax for storing and exchanging data like XML and CSV. And this format is human readable which makes it even better. Example of JSON data:

```
{
     "firstName": "Lokesh",
     "lastName": "Patidar",
     "address": {
         "areaName": "Reva Hostels",
         "city": "Bangalore",
         "state": "Karnataka",
         "postalCode": 560064
     },
     "phoneNumbers": [
         "8769953377",
         "9108271971"
     ]
 }
```

Figure 3 JSON data format

For our purpose we are only interested in the streaming data, the Streaming API of twitter is very useful because twitter does not impose any rate limits on it. So, we can get massive amount of data without exceeding the rate limits.

In this section, we'll implement a custom stream listener by extending the default StreamListener class offered by Tweepy. The core of the streaming logic is implemented in the CustomListener class, which extends StreamListener and overrides two methods: on_data() and on_error(). These are handlers that are triggered when data is coming through and an error is given by the API.

This is twitterStreaming.py which When we run, have to be provided with arguments from the command line. These arguments, separated by a white space, will be the keywords used by the listener to download tweets.

A tweet gained by this technique has a great deal of crude data in it which we could possibly discover valuable for our specific application. It comes as the JSON object however we can change over it into python "dictionary" type with different key-value sets. A rundown of some key-value pairs is given underneath:

- User ID
- Screen name of the user
- Original Text of the tweet
- Presence of hashtags
- Whether it is a re-tweet
- Geo-tag location of the tweet
- Date and time when the tweet was created

Since this is a lot of information we only need original text of the tweet itself which we will extract in next section of design.

Data Pre-processing
After streaming tweets, while going for analysis of sentiment we only need original text of the tweet which is stored in "text" key value of the tweet which is converted to dictionary from JSON object using *json.loads()* function in inbuilt *json* library.

Now that we have got our original text, we will preprocess it before analyzing its polarity towards positive or negative sentiment. We streamed some tweets and stored them into a file, which we will give as command line argument to the *preprocessing.py*.
The main targets of this preprocessing program:

- Lower case - converted the tweets to lower case
- URLs - eliminated all the URLs
- #hashtag - hash tags can give us some useful information, so replacing them with the exact same word without the hash.
- Punctuations and additional white spaces - remove punctuation at the starting and ending of the tweets.
- Feature Reduction (For example: Tokenization, Removing Stop words, Twitter symbols, and Repeated Letters)

An example of tweet in JSON format:


```
{"created_at": "Wed Dec 28 07:44:50 +0000 2016", "id":
814014228000428032, "id_str": "814014228000428032", "text": "RT
@cimpankaj: @Paytmcare @Paytm  If u are not ready to solve this
issue then at least admit that. You are not giving the solution
for my p\u2026", "truncated": false, "entities": {"hashtags":
[], "symbols": [], "user_mentions":

....

"geo": null, "coordinates": null, "place": null, "contributors":
null, "is_quote_status": false, "retweet_count": 1,
"favorite_count": 0, "favorited": false, "retweeted": true,
"possibly_sensitive": false, "lang": "en"}, "is_quote_status":
false, "retweet_count": 1, "favorite_count": 0, "favorited":
false, "retweeted": true, "lang": "en"}
```


Figure 4 Tweet without processed

And its processed text would be:

```
RT @cimpankaj: @Paytmcare @Paytm  If u are not ready to solve
this issue then at least admit that. You are not giving the
solution for my problem
```

Figure 5 Processed tweet

Analysis

The TextBlob package for Python is a convenient way to do a lot of Natural Language Processing (NLP) tasks. Processing (NLP) tasks. For example:

```
from textblob import TextBlob
TextBlob("not a very great movie").sentiment

## Sentiment(polarity=-0.3076923076923077, subjectivity=0.5769230769230769)
```

This tells us that the English phrase "not a very great calculation" has a *polarity* of about -0.3, which means it is somewhat negative, and a subjectivity of about 0.6, which means it is genuinely subjective.

IV. RESULTS AND DISCUSSION

Sentiment analysis has numerous applications and advantages to your business and association. It tends to be utilized to give your business important experiences into how individuals feel about your item image or administration. At the point when connected to online networking channels, it tends to be utilized to distinguish spikes in sentiment, in this way enabling you to recognize potential item promoters or internet based life influencers.

A standout amongst the most all around archived employments of Sentiment Analysis is to get an entire 360 perspective on how your image, item, or organization is seen by your clients and partners. Generally accessible media, similar to item audits and social, can uncover key bits of knowledge about what your business is doing well or off-base. Organizations can likewise utilize sentiment analysis to gauge the effect of another item, advertisement battle, or buyer's reaction to ongoing organization news via web-based networking media.

Sentiment analysis is utilized in business insight to comprehend the emotional reasons why shoppers are or are not reacting to something (ex. for what reason are shoppers purchasing an item? What's their opinion of the client experience? Did client administration bolster live up to their desires?). Sentiment analysis can likewise be utilized in the zones of political theory, humanism, and brain science to investigate patterns, ideological predisposition, conclusions, measure responses, and so forth.

V. CONCLUSION AND FUTURE SCOPE

The assignment of sentiment analysis, particularly in the area of miniaturized scale blogging, is still in the creating stage and a long way from complete. At this moment, we have worked with just the least difficult of calculations among many like Naive Bayes, SVM and so on. we can improve those models by including additional data like closeness of the word with a nullification word.

The best organizations comprehend sentiment of their clients – what individuals are stating, how they're stating it, and what they mean. Sentiment Analysis is the area of understanding these feelings with programming, and it's an unquestionable requirement comprehend for engineers and business pioneers in a cutting edge working environment.

Similarly as with numerous different fields, propels in Deep Learning have brought Sentiment Analysis into the frontal area of forefront calculations.

Future Enhancements

- *Multi-class classification*

Till now, we have just managed double arrangement of tweets, either as positive or negative sentiment. There are numerous tweets, for example, those with URL's which don't have any sentiment, or, are unbiased. These tweets are primarily for imparting some helpful data to individuals, and not really for raising a conclusion. As a piece of my future work, I might want to investigate multi-class arrangement into different dimensions of sentiment, for example, incredibly positive, positive, impartial, negative and very negative.

- *More numeric features*

The numeric highlights that were utilized in this trial incorporate number of negative and positive words, emojis, length of tweets and number of extraordinary characters, for example, outcries, hashtags, etc. The numeric highlights did not yield great precision and gave around 63 percent exactness. Thus, as a piece of my future work on this, I might want to produce more just as more intelligent numeric highlights.

ACKNOWLEDGMENT

We sincerely thank our guide Prof. Raghavendra Nayaka P. for his valuable guidance throughout the process. Last but not the least we would like to thank our Director Dr. Sunil Kumar S Manvi. This paper would not have been possible without his support.

REFERENCES

[1] Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment By Tumasjan https://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/viewFile/1441/1852

[2] Efthymios Kouloumpis, Theresa Wilson and Johanna Moore. Twitter Sentiment Analysis: The Good the Bad and the OMG! In Proceedings of AAAI Conference on Weblogs and Social Media (ICWSM), 2011.

[3] Theresa Wilson, Janyce Wiebe and Paul Hoffmann. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In the Annual Meeting of Association of Computational Linguistics: Human Language Technologies (ACL-HLT), 2005.

