

# Sentimental Analysis of Text and Image Twitter Data for a given location using location information and implementation using open source tools

Sandeep Vaidya<sup>1</sup>, Harshavardhana Doddamani<sup>2</sup> and Bharathi M<sup>3</sup>

<sup>1</sup>Author, [vaidyanet@gmail.com](mailto:vaidyanet@gmail.com)

<sup>2</sup>Author, Assistant Professor, [hdoddamani@gmail.com](mailto:hdoddamani@gmail.com)

<sup>3</sup>Author, Assistant Professor, [bharathigowdal@gmail.com](mailto:bharathigowdal@gmail.com)

<sup>1,2,3</sup> Sri Jagadguru Chandrashekaranaatha Swamiji Institute Of Technology, Chickballapur – 562101, (Karnataka)

**Abstract** – Considering the popularity of Twitter application among the people of all ages, sentimental analysis of twitter data provides extensive information across multiple emerging markets to study the sentiment of specific group of people within the given geographical area. Impact of sentiment and level of abstraction differ from area to area to conclude on the analysis results. Analysis of the data of specific geography helps to overcome these limitations. Sentiment analysis of some of the data may not be applicable for the selected areas, so people belong to specific geography can be selected carefully to conclude on the analysis and later appropriate actions can be implemented to improve the sentiment of the people.

**Index Terms** – Sentimental Analysis, Image analysis, Geo location services.

## I. INTRODUCTION

In recent times, internet is more affordable and reachable to common man. This helps the individual to get connected to the world and provide their opinion on the different topics in which they are interested. Twitter being the popular social platform where people from all age group including common man, celebrities, and political people use it effectively to showcase their thoughts and views on various topics. People who follow their icons get to view these thoughts and update or share their own thoughts on the same subjects.

Many companies are using the social network to study the behaviour of their employees as well as even refer the social network involvement and posts of the potentials candidates before giving them the jobs in the company. If they see any offensive posts or comments from the candidate in the social network or by the existing employee, they may be rejected or even thrown out of the company. Social networks are considered as the mirror image of the individual who hides some emotions in front of people but opens up in the social network.

Social networking sites popularity is rising consistently when the new users are absorbed by various different means. It has become a trend for an individual to be an active in the social network sites where people will continue to respond to each comment of his friends or the people of his interests.

Almost all companies now are thinking of merging or associating their solutions with the social global network sites where it can easily influence people in the social network. They develop interactive communities and tools within the social network sites based on the individual interests or business interests where they can make sure the information about their products can be flown which will influence the users of social network to think information. This would help clinician's faster and more effective diagnosis.

Major uses of the social network study are as follows.

- Create a brand awareness
- Online reputation management tool
- Recruiting
- Learn about new technologies
- Learn about competitor
- Lead generation for the potential prospective customers

Considering the so much uses of social network in the business, even it is very much important to study the sentiment of the people about their own brand. Even the sentiment analysis can be used as the tool in the political circle to study the sentiment of the people on candidate whom they are planning to consider for the next election etc.

This influencing feature can be utilized on the online social network to influence people to help in the marketing of a specific product or an idea. There is a specific type of marketing technique which uses the social media is called "Social Media Marketing". This recently has become a hot topic in the industry. This helps to target the specific area, detect the opportunities and threats in advance and it helps to protect the brand. It helps to improve the reputation and in turn converting this to profit

There is a separate marketing method called the Social Media Marketing has evolved which will be used to study the Social media and sentiment of the social media data to study the market, get the feedback and improve the product and control the market. This way the profitability of the same can be improved. Social media marketing is considered as the one of the best and fastest influencing method.

Even the individual companies can also perform this kind of analysis related to company employees to identify how many people are happy working with them or how many people who are related to that company are happy about them by analyzing the sentiment of the people related that company

*Challenges:* Some of the challenges in social network analysis include, making computer understand each statement and deriving the sentiment is not very easy. Even the statement may use positive words but with negation to change the meaning of whole statement. Statements may include sarcasm makes it difficult to analyze. Even users may post their images to show the sentiment.

Miss diagnosing statements may give completely different picture about the sentiment. So the algorithm used shall be as accurate as possible. Minimum of 80% accuracy will help us to provide better analysis and sentiment index.

Sentiment analysis can be done on various different types of data. It can be done on the any organizations social gathering website or government feedback website or based on the product review feedback etc. Social media like twitter can be an easy option and we can get plenty of data with different sets of users, different regions of users and users with completely different culture.

Sentiment analysis can be done based on various methods:

- Lexicon based method uses the ranking based analysis
- NLP based analysis uses the natural language processing to understand the meaning of the statement to derive the sentiment.
- Sentiment Analysis of Images in combination of text to derive more meaningful information

There are various open source tools/services to achieve the sentiment analysis of the text and images.

#### **A. *Sentiment Analysis of Text***

NLP based sentiment analysis tries to understand the meaning of the statement to decide on whether the statement is positive or negative. This requires the algorithm which will continue to learn from the statements and the accuracy of the analysis will improve as the learning progress

Stanford University's **Neural Analysis of Sentiment (NaSent)** algorithm is implemented based on the recursive neural network that builds or operates on top of grammatical structures. This algorithm uses the trained model to identify the sentiment of the statement.

This module is designed to take the text as input and return the sentiment as the output as positive, negative or neutral

Library provided by the Stanford University can be used for this purpose. Link <http://nlp.stanford.edu/sentiment/> provides details on how the library can be used and also the functional details of the algorithm.

#### **B. *Sentiment Analysis of Image Twitter Data***

Image data also can be streamed with the text data. These images are downloaded as the file in the solution which can be passed as the input to the Image sentiment analysis engine to identify the sentiment of images. Sentiment of Images and sentiment of the text together provide the more accurate sentiment of the twitter tweet.

Indico online API provide the service for sentiment analysis of images. Each Images data can be converted into the format supported by the Indico before passing it to get the sentiment of images.

Using of Indico requires the application to be registered with Indico to get the API KEY. Later the API key to be passed with each request to get the sentiment of the image.

### C. Design of the integrated platform

Integrated platform can be designed which will stream the twitter data and process the text and Image data using the text analysis engine and Image analysis engine. Once the analysed data is captured, same can be displayed using the visualization tools to display the analysis results.

Refer fig 1 for the integrated platform architecture diagram below.

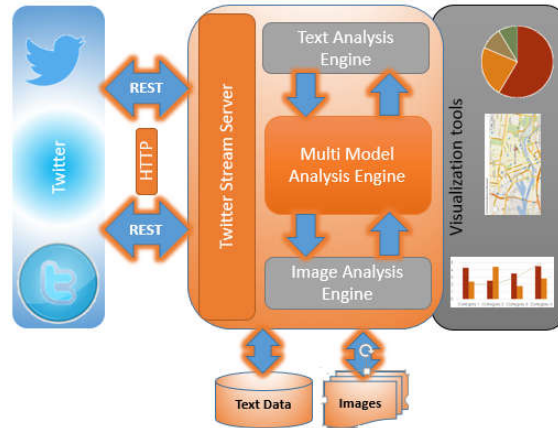


Fig. 1: Point mapping

*Twitter Stream Server:* Stream server Module will have two components. Server modules will start the streaming thread. This streaming thread will connect to the twitter platform and download the Twitter data. Similar way multiple threads can be created for downloading the data from different social network platforms if we want to integrate more social platforms like Facebook etc. Twitter exposes the REST based API's to stream the twitter data by opening the handle to the public stream. Using the properties of the URL i.e. `include_entities=1`, we can even get the images existing in the tweets as we are focussed on the analysing the images as well. Each read line will download the one complete tweet which will have the complete details of the tweet like tweet text, user id, tweet location, image URL etc.

*Parsing Twitter Data:* Twitter return the data in either the JSON format or the XML format. JSON is more friendly and easy to understand. JSON parser can be used to extract individual elements from the twitter data like tweet text, tweet location, user name and any image URL attached with the tweet. Tweet text, location, date, user details can be stored into the database. Using the image URL, image can be downloaded using any of the HTTP APIs on the selected folder.

Once the tweet is extracted and image is downloaded, both of them are separately analysed using the respective engines. Tweets are stored in database with the respective sentiment of the statement.

The *geocode* parameter in the Twitter API helps to capture the tweets from the specific location. This parameter help us to focus on the given geo location. Geo location can be passed to this API using the bounding box of the tweet.

*Geo Code Challenges:* There is a major challenge in getting the geo code of the tweet. Most of the twitter users at least in country like India would not have enabled the location services in mobiles. Due to this the location information will not get captured with the tweet. Even in other countries, users who want to tweet without giving location information will not enable the location services or do not share the location information with the twitter application. If this is the case there is no way we can get the tweet location.

*Work Around for Geo Location:* Work around to get the tweet location is by using the user address or country and city. We can use the google location API and pass these address details to get the bounding box. This bounding box can be considered against the tweet location. This does not mean the location information got using this work around is accurate, but we assume it is near to accurate.

*Text Analysis Engine:* Text analysis engine is again created as a separate thread which will take the extracted tweet text to apply the sentiment analysis using the Stanford University's NaSent library as given in the section "Sentiment Analysis of Text"

*Image Analysis Engine:* Image analysis engine uses the Indico online API to analyse the sentiment of images. Downloaded image data need to be converted in the format Indico API expects and pass the complete image as the byte array to get the sentiment. Once the sentiment of the image is received, it can be saved into the database with the respective twitter tweet id. Either we can use the tweet text sentiment and image sentiment together for better accuracy or in separation.

Once the sentiment analysis data is captured in the database, we just need to draw the data using visualization tools to show the analysis results.

*Visualization Tools:* Visualization tools shows the data in the different analytical tools. Based on the geolocation, different charts can be drawn to study the sentiment and even compare the data across the geo locations. Fig 2 shows the sample charts in the form of Pie charts of sentiment of the particular topic.

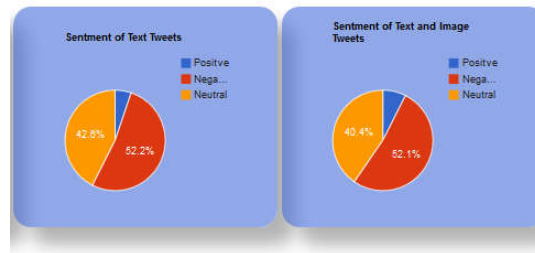


Fig. 2: Visualization tools

Google provide the free JavaScript based free visualization library to draw various different graphs.

*Tweet Location Marking:* Tweet location and the tweet text can be marked on the map to even show the location of the tweet. Just by moving the mouse button over the tweet mark, it shows the tweet text. This can be achieved using the google API which will help us to draw the marks on the map for the location of the tweet and even associate the tweet text with each marker.

Fig 3 shows the sample tweet locations drawn over the map.



Fig. 3: Tweet Locations

## RESULTS, DISCUSSION and CONCLUSION

Sentiment analysis of text produced accurate results. However the sentiment analysis of image still needs additional work. Images with facial structure not clear means, you will not get the accurate sentiment.

- Twitter tweets are analyzed for sentiment using NLP and correct sentiment is recorded
- Images and twitter tweets are extracted from the twitter server and stored into the database with sentiment parameter for future use
- Map displayed with correct set of tweets
- Charts with the correct analysis is displayed

Some of the parameters studied related to the results are:

*Speed/Performance:* Performance of the solution is good for the text sentiment analysis. Image sentiment analysis performance is not that great.

*Accuracy*: Accuracy is good for text analysis, but the Image sentiment analysis still needs improvement. Image sentiment also works mostly for the images with human faces. Even sentiment to be derived from any images like images of crime etc.

## REFERENCES

- [1] <http://nlp.stanford.edu/sentiment/> - NLP algorithm and library
- [2] <http://dimplejs.org/> - Using D3 to visualize the data
- [3] Google Maps API - <https://developers.google.com/maps/documentation/javascript/>
- [4] Google visualization library - <https://developers.google.com/chart/>
- [5] Twitter Stream APIs - <https://dev.twitter.com/streaming/overview/request-parameters>
- [6] <http://twitter4j.org/> - Twitter Streaming APIs
- [7] Struts UI Frame work - <http://www.java4s.com/struts-tutorials/>
- [8] Image Emotion recognizer - <https://github.com/mpillar/java-emotion-recognizer>
- [8] Image Sentiment analysis - <http://www.l3s.de/~siersdorfer/sources/2010/mm10-siersdorfer.pdf>
- [9] Stanford recursive neural network based sentiment analysis - [http://nlp.stanford.edu/~socherr/EMNLP2013\\_RNTN.pdf](http://nlp.stanford.edu/~socherr/EMNLP2013_RNTN.pdf)
- [10] <https://indico.io/> - Machine learning for text and image data into human insight