

SENTIMENT ANALYSIS OF TWEETS

Sharmila Patil¹, Ms.Nalini Jain², Mr.Yasar Shaikh³, Ms.Supraja Sakhare⁴

¹ Assistant professor, Information Technology, WIT, Solapur, Maharashtra, India

² Student, Information Technology, WIT, Solapur, Maharashtra, India

³ Student, Information Technology, WIT, Solapur, Maharashtra, India

⁴ Student, Information Technology, WIT, Solapur, Maharashtra, India

Abstract— Social media analytics is an influential means for uncovering human sentiment across countless online sources. The analytics facilitate to various fields to identify sentiment and identify trend and inclination of people. Sentiment analysis method involved observation and monitoring different public sentiment and qualitatively exploring the underlying content.

In proposed method twitter data get scrutinized for the sentiment analysis. This is implementing polarity classification over Tweet offered on multiple keys to monitor the feelings/emotions of general public towards bring up issues. These posts are classified as positive, negative or neutral. The project focuses on addressing polarity classification for topic featured in tweet by linking semantic association between the features and polarity words.

Key Words: Social media, tweet, sentiment analysis, polarity, tweet monitor.

I INTRODUCTION

Social media is easily accessible an online tool where content, views, perception, opinions, and media can be shared. Some people create content, while others lurk, observe, or disseminate content. At its core, social media is about relationships and connections between people and organizations (Nair 2011)[1]. Apparently wisdom of the crowd, which is documented on the web, plays a key role in our major decisions.

Twitter is one of the social media, also called a micro blogging service launched in 2006 with more than 20 million unique monthly visitors. On Twitter, every user can publish short messages with up to 140 characters, so-called “tweets”, which are visible on a public message board of the website or through third-party applications. There are more than one million messages per hour posted on twitter (Andranik Tumasjan et al 2010)[2]

Recently, the exponential growth of Twitter has started to draw the attention of researchers from various disciplines. There are several streams of research investigating the role of Twitter in social media, product marketing, and project management. One stream of research concentrates on understanding micro blogging usage and community structures (e.g., Honeycutt and Herring 2009; Huber man, Romero, and Wu 2008; Java et al. 2007) [3].

We use particularly Twitter as Micro blogging platforms as it is used by different people to express their opinion about different topics, thus it is a valuable source of people’s opinions. Twitter contains an enormous number of text posts and it grows every day. The collected corpus can be arbitrarily large. Twitter’s audience varies from regular users to celebrities, company representatives, politicians and even country presidents. Therefore, it is possible to collect text posts of users from different social and interests groups. Sentiment analysis has been an important topic for data mining, while the prevailing of social networking, more and more tweet analysis research focuses on social networking.

In this paper we introduce a sentiment analysis tool, it comprises three functions: sentiment analysis among Twitter tweets, finding positive, negative and neutral tweets from information resources. This tool focuses on analyzing tweets from those media sites.

The aim of this project is fourfold. First, we gather opinions from twitter. Tweets which contain the specified keywords and hash tags are captured. These tweets along with its metadata are stored in

individual xml files. Then the stored data is ingested to HDFS. The resultant data is then subjected to Hadoop for sentiment analysis.

The project will enable us to do different types of Sentiment analysis on trending topics and display those results by various visualization techniques. The successful use of social media has established Twitter, Face book.

II. LITERATURE SURVEY

To get an idea from content posted by users, a “sentiment analysis” was obtained. A sentiment analysis also referred as opinion mining is one of the advanced applications of NLP that focuses on content in terms of their polarity (positive, negative, and neutral). It is collection of human intelligence and electronic intelligence for mining the text and classifying user sentiments, likes, dislikes and wishes. It is the important aspect for capturing public opinion about different domains. Businesses, academics, and journalists are using sentiment analysis to tap into this social media buzz to gain a better understanding of consumer preferences, market trends and brand awareness. From the consumers’ point of view, using sentiment analysis can leverage the way for research on products and services before making a purchase (Go et al. 2009)[4]

Currently, a wide range of tools have been developed that uses sophisticated computing paradigms, which include machine learning algorithms to find insights and patterns for giving text. O’Connor et al. (2010) [5] use a lexicon projection strategy yielding predictions which significantly correlate with polls regarding ratings of Obama. While it is clear that deeper linguistic analysis should be performed for better results (Pang and Lee, 2008).

Spurious et al. proposed the Twitter follower graph to improve sentiment classification and constructed a graph that has users, tweets, word unigrams, word bigrams, hash tags, and emoticons as its nodes which are connected based on the link existence among them. They then applied a label propagation method where sentiment labels were propagated from a small set of nodes seeded with some initial label information throughout the graph. They claimed that their label propagation method outperforms MaxEnt (Maximum Entropy Modeling) trained from noisy labels and obtained an accuracy of 84.7% on the subset of the twitter sentiment test set.

Conover et al. examined the retweet network, where users are connected if one re-tweet tweets produced by another, and the mention network, where users are connected if one has mentioned another in a tweet, of 250,000 political tweets during the six weeks prior to the 2010 U.S. midterm elections. They found that the retweet network exhibits a highly modular structure, with users being separated into two communities corresponding to political left and right.

In addition to sentiment analysis for English and other European languages, sentiment analysis is applied on various Indian languages like Bengali, Hindi, Telugu and Malayalam also.

General Inquirer system developed by (Stone, 1966) Philip Stones, was the first milestone for extracting textual sentiment. It is based on the manual database containing set of positive or negative orientations and the input words are compared with database to identify their class such as positive, negative, feel, pleasure.

(Turney, 2002) For the classification of positive and negative opinion, Peter Turney proposed the idea of Thumbs Up and Thumbs Down. For better problem formalization, there was the necessity of an automated system, which could be employed for electronic documents. For consecutive words and their polarity, Turney came up with an algorithm to extract Point wise Mutual Information(PMI).Experiments were conducted on movie review corpus and polarity is referred to as "thumbs up" for positive and "thumbs down" for negative.

Barbosa and Feng (2010 use polarity predictions from three websites as noisy labels to train a model and use 1000 manually labeled tweets for tuning and another 1000 manually labeled tweets for testing. They propose the use of syntax features of tweets like retweet, hash tags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words.

III. METHODOLOGY

We collected political tweets, which were published on Twitter's public message board on February 8 and February 9, 2015. The model is going to receive live tweets from twitter. After several steps search each tweet for words positive or negative polarity. Along with keywords the tweets in dataset may contains positive emoticons, such as ☺ and the other half contains negative emoticons such as ☹. This is binary classification task with two classes of sentiment polarity : positive and negative.

Some major commercial features of Twitter can be explained as follows:

- a) **Hashtag:** meta-tag beginning with “#” that is designed to help others find a post. This feature seems to be invented by Twitter users in early 2008 (Huang et al. 2010)
- b) **@ Symbol:** In contrast with a hashtag, the @ symbol is used to address a tweet to another registered Twitter user (Honey, Herring 2009).

The Front End portal is used to display the detailed dynamic content to the user

IV. WORKING MODEL

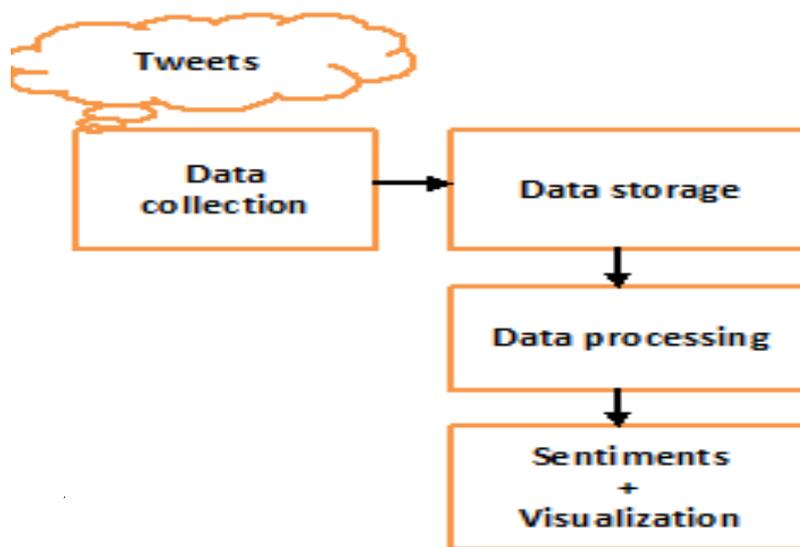


Fig 1. Working model

V. ARCHITECTURE

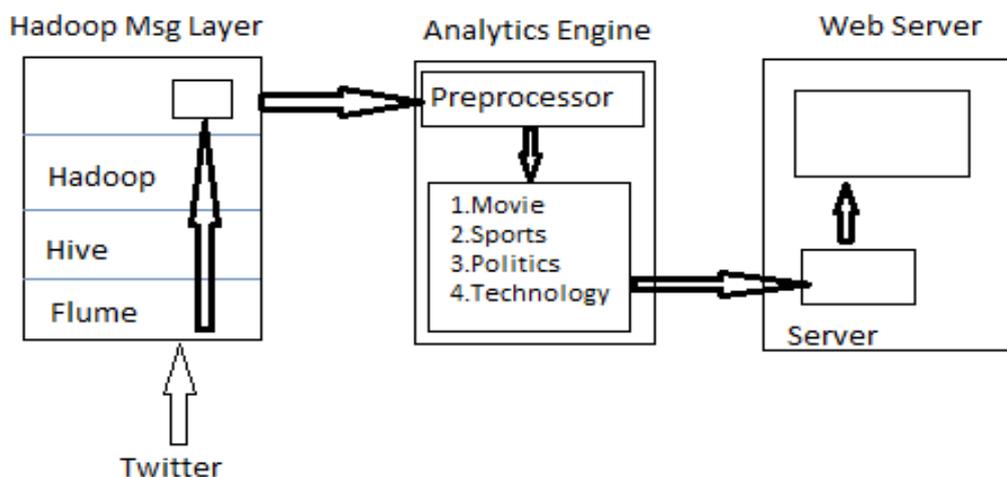


Fig 2. Architecture of analysis model

VI. IMPLEMENTATION

6.1: Twitter Crawler –

- First the twitter stream is crawled for specific tweets.
- The tweets are filtered according to keywords.
- These tweets are stored in individual xml files.
- The xml files obtained from the first step are stored in the database

6.2: Processing-

- The Tweet and LatLong from the individual XML files is extracted.
- Then the sentiment is found out, translation is done and the result is stored in corresponding XML File.

6.3: Server

- The ingested files are range indexed on attributes for which faceting is to be done.
- Range indexes are also specified for the latitude and longitude elements.
- The Application builder is used to build the application and build facets. And widgets are used to aggregate some of its aspects.
- The final sentiment is calculated based on the total positive, negative count

6.4: Visualization

- By using different visualization techniques we will be displaying the final results in graphical formats.
- Visualizations make easy understanding.

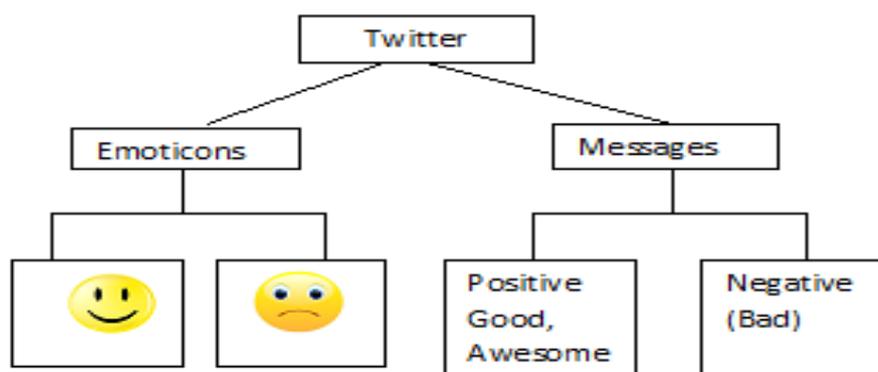


Fig 3. Polarity Calculation

This application will spot keywords posted on tweets. Keywords are spotted for particular words e.g. "good #movie #ChakDeIndia". The word "good" makes the polarity value as the '+1' i.e. the increment in the polarity values. A negative review will decrease the polarity values. The results will be shown using rating meter. The main advantage is that the application is dynamic in context of words. Like nowadays there are new words like "LOL, ROFL, OSSAM", which is hard to categorize with the help of NLP.

The application is quite faster than the NLP approach, because the NLP takes more time to analyze the context of speech. We are using the EMOTICONS, which also can be increased as we have separate word list, we can add in that. The drawback of this application is that Tweets containing multimedia will not be processed in this model and also the tweets having different fonts/languages except English is not processed in this model. But this can be overcome by use of EMOTICONS, because it is same for all languages. And we can have separate word list according to that. But ultimately increasing word files results in slower performance of the application.

VII. CONCLUSION

Twitter can indeed be used as a platform for different types of analytics. The mere number of opinions reflects views of people preferences and comes close to predictions while the sentiment of messages closely corresponds to moods of people .Different types of visualization techniques can be used to understand the results effectively.

REFERENCES

- [1] Nair, M.”Understanding and measuring the value of social media”. Wiley periodicals, pp.45-51.
- [2] Andranik Tumasjan, Timm O. Sprenger, Philipp G. Sandner, Isabell M. Welpé “Predicting Election s with Twitter: What 140 Characters Reveal about Political Sentiment, 2011.
- [3] Das, S.R. & Chen, M.Y. “Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web”, 2007
- [4] Go, A., Bhayani, R. & Huang, L., “Twitter sentiment classification using distant supervision”. CS224N Project,2009.
- [5] Brendan O’Connor, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. “from tweets to polls: Linking text sentiment to public opinion time series”,2010.

