Reviewing Sentiment Analysis and Opinion Mining of Social Media

Sheenu Choudhary, Er. Dipika Arora

M.Tech Student, Dept. of CSE, RPIIT, Karnal, Kurukshetra University, India

ABSTRACT: Sentiment Analysis is the widest term which has been adapted in Machine Learning. The word "Sentiment" refers to a person's view or opinion and analysing the same views is known as Sentiment Analysis. In machine learning language it can be defined as one of the data mining technique which identify, extract or filter the user's review or surveys. This paper is focusing on the domain of Sentimental Analysis for twitter posts, news, blog and some other text articles. The main objective here is to check whether certain piece of information is positive, negative or neutral. Opinion Mining is one of the similar terms used for Sentimental Analysis because it derives the opinion of the users through their tweets. We cannot deny the fact that sentiment analysis is very complex, especially in Social Media World due to its huge scope and length. So to identify the sentiments from text there are two different techniques i.e. **Symbolic technique** and **Machine technique.** Machine techniques are quite easier and fast as compare to Symbolic techniques.

KEYWORDS: Opinion Mining, Sentiment Analysis, Feature Based Sentiment Analysis, Social Media.

I. INTRODUCTION

Sentiment Analysis is generally the procedure of discovering the meaningful subjective information using natural language processing, text analysis and computational linguistics techniques. The main aim of this approach is to identify and extract that particular information which defines the attitude of the writer or speaker towards a specific idea or topic. The attitude can be either in terms of positive, negative or neutral.A common use case for this technology is to discover how people think about a specific area of topic. This is one of the major fields which researchers are targeting nowadays. News and blogs are usually very fine platforms for sentiment analysis, because the users are allowed to express their thoughts and ideas with full freedom. The processing system includes a sentiment identification phase in the beginning wherein it detects the different entities from the available twitter posts and relates the expressed opinions with its relevant entity. Then, the second stage includes sentiment aggregation and scoring where every entity is scored related to others in one class. At last the major area of focus is to detect the semantic group of words in the initial stage, using word Net. However, there are several challenges however in sentiment analysis. Some of them are- object identification, feature extraction and synonyms grouping, opinion orientation classification and integration of all the above.

II. LITERATURE SURVEY

According to [1], the authors have proposed an adaptable approach that automatically extracts different set of patterns of textual information which has similar meanings and sentiments in tweets.In previous approach ,sentiments pattern extraction were depend on fixed sets of templates and patterns .Because of fixed sets of patterns no deep analysis was required in tweets. But then in their approach where sentiment analysis is not based on fixed no of patterns deep analysis is required. So they have evaluated their approach with tweet- and entity-level sentiment analysis tasks by capturing the semantic patterns as classification features in both tasks.

They have used 9 Twitter datasets in their evaluation and then compare the performance of patterns against 6 state-of-the-art baselines. After comparing the performance Result shows that their patterns continuously perform better than all others.

In this [2], Micro blogging is very efficient and popular communication tool among Internet users. Every day billions of users share their opinions, different views and their perception regarding life. In another words micro blogging web-sites are good sources and collections of information for this process. As micro blogging has been launched recently, so there are quite few research works that have been devoted to this particular domain. They are basically focusing on Twitter, the social networking domain and the most widely used micro blogging platform to perform sentiment analysis. Using the collection, they have tried to develop an opinion classifiers with the help of which they can catch good, bad and unemotional opinion for a report.

According to this paper[3], with the growing of shortly written blogs on the web, people started share their ideas or views on a large collection of topics on Twitter.Therefore, Sentiment analysis on articles like products, organizations, people, etc becomes a fast and efficient way of calculating public opinion for business marketing or social studies. Because of Twitter's unique characteristics some new problems comes for present sentiment analysis methods, which is basically focused on huge biased collection . In this paper, they come up with a new entity-level sentiment analysis method for Twitter. This method firstly uses a lexicon based approach to execute sentiment analysis of entity level. And In this paper, they proposed a different method to face the problems. Firstly we applied an augmented lexicon-based method to the Twitter data to perform sentiment analysis.

Proposed Methodology And Discussion
There are two approaches (categories of Classifiers) for sentiment analysis
- Lexicon Based Method
- Machine Learning Method

First of all data need to be pre-processed for Sentiment Analysis as mentioned above. Data preprocessing steps are common in both the techniques. After pre-processing any one of the two techniques can be used.

A. Lexicon Based Method:
In this method polarity of each word in BOG (bag of words) is calculated. Then average score of all the unigrams is calculated with the help of already present polarity dictionary (eg. WordNet).

$$\text{Score AVG} = \frac{1}{m}\sum_{i=1}^{m} W_i$$

That score give the polarity of Twitter Posts. Polarity of each word in unigram can be calculated by using WordNet and SentiWordNet.

B.Machine learning Method:
   In this method Training and Testing of tweets are used. Manually construct training data set and classify it as positive, Negative or Neutral. Then test is done on testing data for verifying the accuracy of model. Then this model can be used to classify the new incoming tweets. The most popular learning algorithms for text classification are: Result-
- Support Vector Machine(SVM)
- Naive Bayes
- Decision Trees (C 4.5)
- Maximum Entropy Algorithm
- AdaBoost
- Regression Method (Logistic Regression LR)
- J48
- Simple Cart
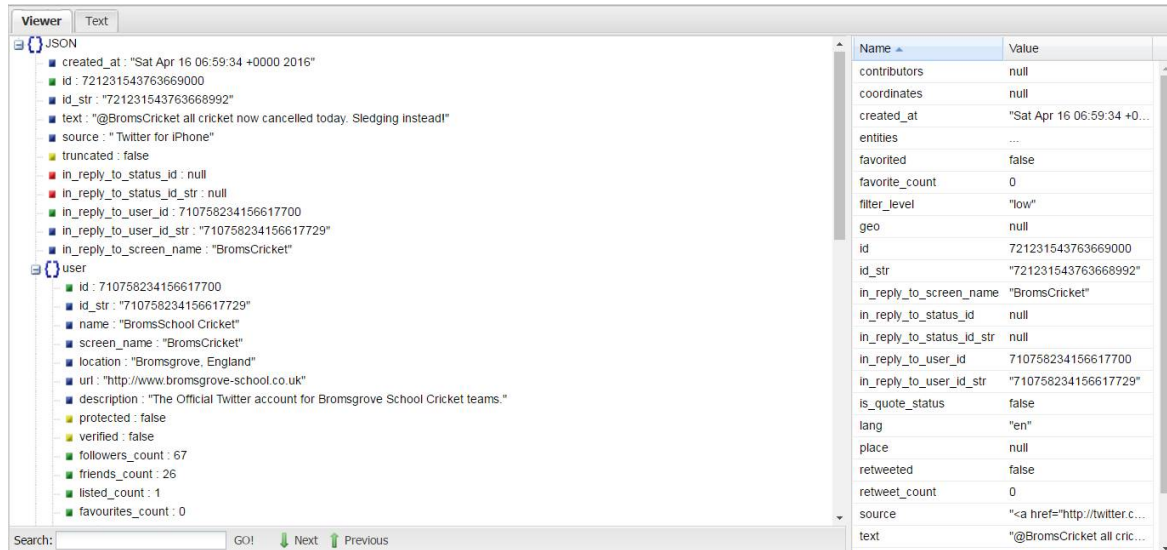- Random Tree

III. RESULTS AND ANALYSIS

Figure 1: Retrieved JSON from the Twitter API visualized through
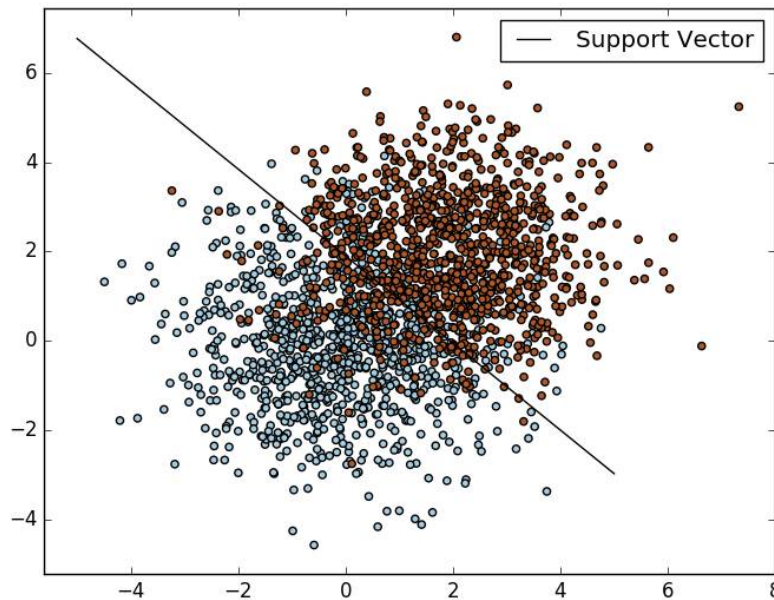
Figure 2: Decision Boundary for Positive (red) and Negative (blue) tweets using Support Vector Machines
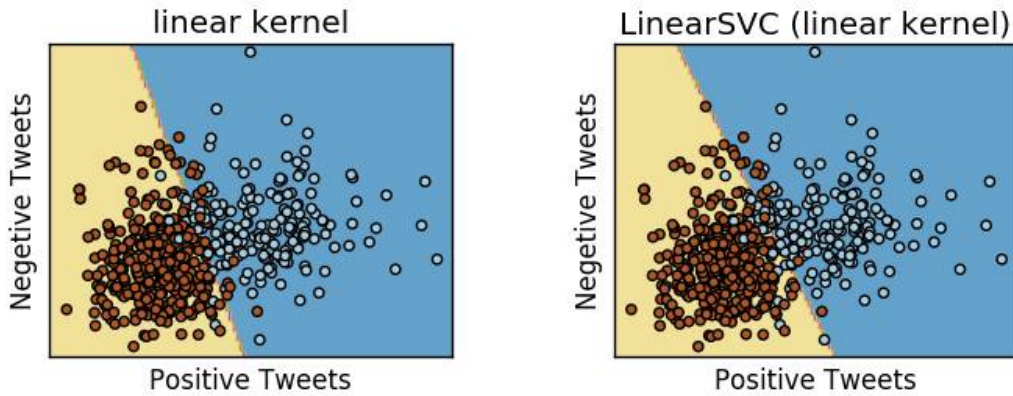
Figure 3: Comparison of Classification Boundary for Positive (red) and Negative (blue) tweets using Linear and Linear SVC Classifier
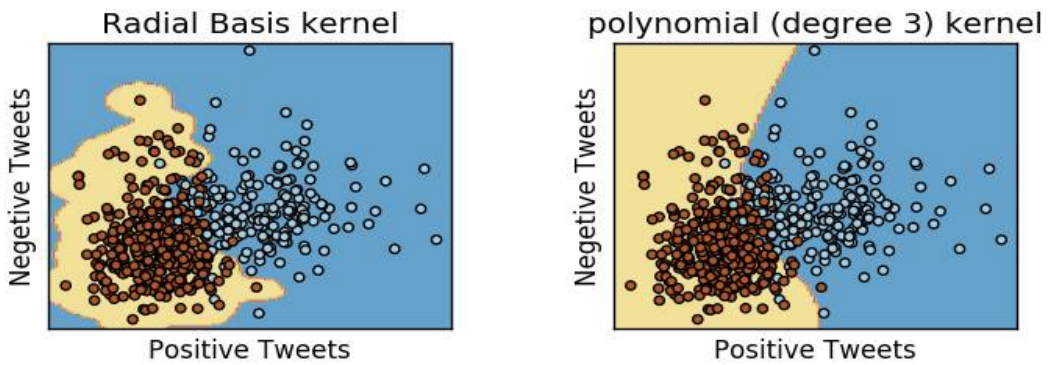

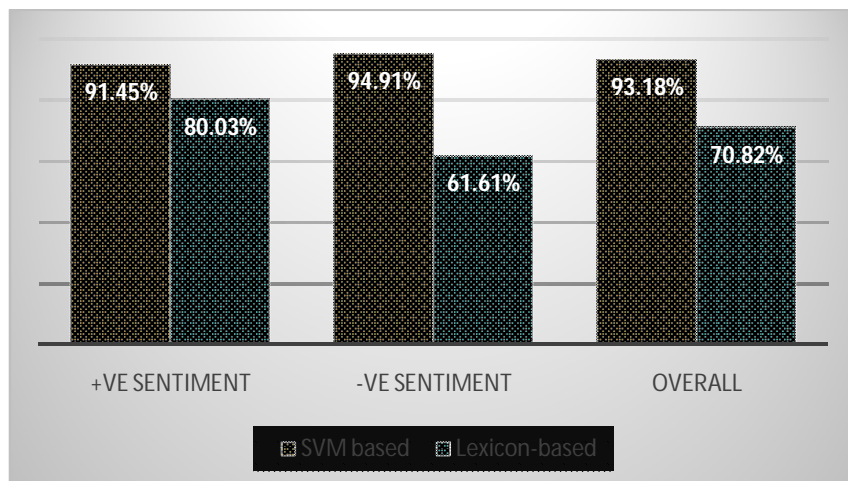**Figure 4: Comparison of Classification Boundary for Positive (red) and Negative (blue) tweets using Radial Basis kernel and Polynomial Kernel**

IV. CONCLUSION

Sentiment research is speech processing task that uses a computational way to recognize opinionated content and categorize it as affirmative or negative. The unstructured textual data on the Web frequently carries expression of opinions of users. Emotion research attempts to recognize the expressions of opinion and mood of writers. An easy emotion research algorithm endeavors to categorize a document as 'positive' or 'negative', established on the opinion expressed in it. In upcoming works we will familiarize an way to emotion research that uses prop vector mechanisms (SVMs) to hold jointly varied origins of potentially pertinent data, encompassing countless favorability measures for phrases and adjectives and, whereas obtainable, vision of the case of the text. Lexicon established classifier will be contrasted alongside SVM established Classifier.

REFERENCES

[1] Hassan Saif , Yulan He, Miriam Fernandez, and Harith Alani. "Semantic patterns for sentiment analysis of Twitter." In The Semantic Web–ISWC 2014, pp. 324-340. Springer International Publishing, 2014.
[2] Pak, Alexander, and Patrick Paroubek. "Twitter as a Corpus for Sentiment Analysis and Opinion Mining." In LREc, vol. 10, pp. 1320-1326. 2010.
[3] Li Dong, Furu Wei, Ming Zhou, and Ke Xu. "Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis." In Twenty-Eighth AAAI Conference on Artificial Intelligence. 2014.