RESEARCH ARTICLE OPEN ACCESS

Sentimental Analysis of Movies Tweets with Different Analyzer

Manish Gupta [2], Pankaj Sharma [2]
M.Tech. Scholar [1], Assistant Professor & HOD [2]
Department of Computer Science & Engineering
MDU Rohtak

ABSTRACT
Currently, Social networking site data is part of our life. Social networks, e-commerce websites, and etc., are main sources generating huge text data by users. In these websites, opinions or text present on can be explored for getting sentiment or opinion of users with sentiment calculations. Hence, it becomes essential to make opinion mining of social media data (Big data) to make predictions. To deal with these challenges, the contribution of this paper includes the different approaches to calculate and compare opinion of users for movies by applying Textblob, SentiWordNet and Word Sense Disambiguation (WSD) sentiment analyzers to calculate the polarity of tweets.
Keywords:- Big Data, Opinion Mining, Machine Learning, Twitter, WordNet, Textblob, SentiWordNet and Word Sense Disambiguation (WSD)

I. INTRODUCTION

In recent years, a huge number of people have been attracted to social-networking platforms like Facebook, Twitter and Instagram. So with the rapid increase of World Wide Web[1], people often express their sentiments over internet through social media, blogs, rating and reviews. Most of them use social sites to express their emotions, beliefs or opinions about things, places or personalities.

Moreover, social media provides an opportunity for businesses by giving a platform to connect with their customers for advertising. Businesses mostly depend upon users of social media site's content to a great extent for checking the opinion of users. For e.g. if someone wants to watch a movie, then they firstly look up its reviews online, discuss about it on social media before taking a decision. The amount of content generated by users is too vast for a normal user to analyze. So there is a requirement to analyses the opinion of users ,with different sentiment analysis techniques.

There are various Textual Information retrieval system mainly use factual data for processing, searching or analyzing. Facts have an objective component but, there are some other textual contents which express subjective characteristics. For Sentiment Analysis (SA),core features are mainly opinions, sentiments, appraisals, attitudes, and emotions. due to the huge growth of available information on online sources like Social media, it makes many challenges to develop new applications. For example, recommendation system can predict movies to user by taking opinion of their friend's opinion such as positive or negative of the movie by making use of Opinion Mining.

Machine leaning algorithms are very often helpful to classify and predict whether a document represents positive or negative sentiment. Machine learning is categorized in two types known as supervised and unsupervised machine learning algorithms [2]. Supervised algorithm uses a labeled dataset where each document of training set is labeled with appropriate sentiment. Whereas, unsupervised learning include unlabeled dataset where text is not labeled with appropriate sentiments. This study mainly concerns with supervised learning techniques on a labeled dataset.

Sentiment analysis is usually implemented on four levels namely word level, sentence level, document level and aspect level [3].

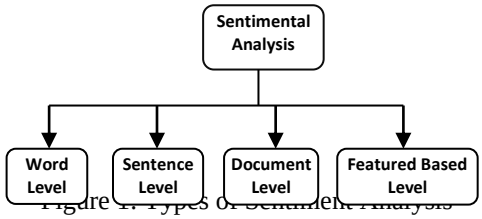


Figure 1: Types of Sentiment Analysis

I. **Word level:** Word level sentiment analysis utilizes adjectives and adverbs that define the sentiment of each word. Two methods that comment on sentiment at word level are :
   i. Dictionary-based approaches
   ii. Corpus-based approaches

II. **Sentence or phrase level**: In this sentiment analysis, each sentence is categorized as either positive, or negative and may be neutral as well. If a sentence turns out to be neutral, it means there is no opinion. All the sentences can then be combined to find polarity of a paragraph or even complete document.

III. **Document level:** In this, whole document is classified as either positive or negative. Respective words and sentences are checked for sentiments

and are subsequently combined to find the sentiment polarity of the complete document.

IV. **Feature-level or aspect-level:** It helps to analyze what people are trying to suggest. Feature level tries to extract sentiment from the opinion directly.

Recent studies have proven that with Twitter it is possible to get people's insight from their profiles in contrast to traditional ways of obtaining information about perceptions. Furthermore, authors of [4] proposed an algorithm for exploiting the emotions from tweets while considering a large scale of data for sentiment analysis. To identify social communities with influential impact, a novel method was proposed by [5] and implemented by assigning metric value to each of the user's emotional posts. Subsequently, the contribution of this paper includes the analysis of movies sentiments gathered from Twitter profiles, with various sentiment analyzers. In addition, this paper presents the comparisons of results obtained from each analyzer. In our analysis comparison of different sentiment analyzers are done with the different classifiers. Sentiment prediction is done on Twitter data to show which technique has a better capability.

II. LITERATURE SURVEY

The analysis of hollywood movies was done on the basis of real-time Twitter data [6], extracted from Twitter by using Twitter-streaming application programming interface (API) [7].
Two sentiment analyzers named SentiWordNet [8] and WordNet [9] were used to find positive and negative scores. To add accuracy to the model, negation handling [10] and word sequence disambiguation (WSD) were used. Twitter streaming API was also used to gather data by the authors of [11] for the prediction of the Indonesian presidential elections. The aim was to use Twitter data to understand public opinion. For this purpose, after the collection of data, the study performed automatic buzzer detection to remove unnecessary tweets and then analyzed the tweets sentimentally by breaking each tweet into several sub-tweets. After that, it calculated sentiment polarity and, to predict opinion outcome, used positive tweets associated with movie, Twitter-based prediction was 0.61% better than the same type of surveys conducted traditionally.

III. METHODOLGY

In this paper we have explained a process from the collection, sentiment analysis, and classification of Twitter opinions. We collects the tweets of particular movies & considered tweets that were posted by users in the form of hashtags to express their opinions about current movies opinion. We then pre-processed the dataset by :

- removing stop words,
- removing all URLs (e.g. www.abc.com), hash tags (e.g. #movie), targets (@username),
- Correct the spellings; sequence of repeated characters is to be handled.
- Remove all punctuations ,symbols, numbers

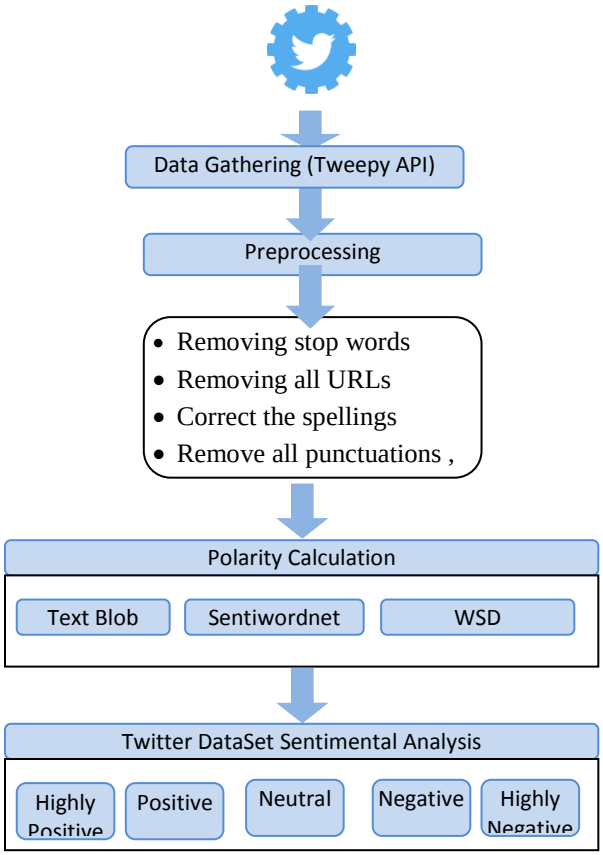
After preprocessing ,Polarity and subjectivity were calculated using three different libraries, SentiWordNet, W-WSD and TextBlob. The following steps explain the tools and techniques that used us in sentiment analysis. Furthermore, sentiment-analysis framework is explained in Figure 1

A- Data Gathering

To gather tweets from twitter based on our movies keywords to view the opinion of the public about these movies. For this, we used Tweepy API [7,12]. We have created an account on Tweepy API linked to our Twitter account. To retrieve the tweets, Tweepy API accepts parameters and provides the Twitter account's data in return. Retrieved tweets, from Twitter accounts, were further used for opinion mining. under the following fields: twitter_id, hashtag, tweet_created, user_id, screen_name, tweet_text, retweet_count, follower_count, and favourite_count of each tweet. We select hashtags that were trending on Twitter, which represents public view for the movies.

```
def DownloadData(self):
    # authenticating

    consumerKey = '[redacted]'
    consumerSecret = '[redacted]'
    accessToken = [redacted]
    accessTokenSecret = '[redacted]'
    auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
    auth.set_access_token(accessToken, accessTokenSecret)
    api = tweepy.API(auth)

    # input for term to be searched and how many tweets to
      search
```

```
searchTerm = input("Enter Keyword/Tag to search about: ")
NoOfTerms = int(input("Enter how many tweets to search: "))

# searching for tweets
self.tweets = tweepy.Cursor(api.search, q=searchTerm, lang =

"en").items(NoOfTerms)
```



B. Data Pre-processing

In the pre-processing step, we removed the irrelevant Twitter data. First, we just kept tweets that were in English.
Furthermore, we pre-processed the dataset by :

- Removing stop words,
- Removing all URLs (e.g. www.abc.com), hash tags (e.g. #movie), targets (@username),
- Correct the spellings; sequence of repeated characters is to be handled.
- Remove all punctuations ,symbols, numbers (e.g. !"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~).

we pre-processed from tweets, because all the above conditions will get information that was not a requirement for sentiment analysis in our approach.

C. Polarity Calculation and Sentiment Analysis

From social media platforms we can collect valuable information of sentiment analysis by detecting emotions or opinions from a large volume of data present in unstructured format. Sentiment analysis includes five polarity classes, which are strong negative, negative, neutral, positive, highly positive. The polarity of each tweet
is determined by assigning a score from -1 to 1 based on the words used, where a negative score from -1 to -0.5 means a highly negative sentiment, negative score from -0.5 to 0 means a negative sentiment, a positive score from 0 to 0.5 means a positive sentiment, a positive score from 0.5 to 1 means a highly positive sentiment while the zero value is considered a neutral sentiment.

We used Textblob, SentiWordNet and Word Sense Disambiguation (WSD) sentiment analyzers for detecting the polarity of movie reviews, and to give a clear view of the most accurate analyzer for the polarity calculator. Textblob [8] is a Python (2 and 3) library for processing textual data. Natural language processing (NLP) tasks such as stopword removal, url removal, noun phrase extraction, part-of-speech tagging, selling correction ,sentiment analysis, classification, translation, and more are provided in a simple API. Similarly, we used SentiWordNet [9], which is a publicly available analyzer of the English language that contains opinions extracted from a wordnet database. In addition to that, W-WSD [11] has the ability to detect the correct word sense within a specified context. Baseline words/Unigrams are clear representatives for the calculation of polarity. Consequently, we used Unigram data in WSD. Despite admitting uncertain observations, WSD is far better than other natural-language processing context recognition functions.

IV. RESULT AND DISCUSSION

This section contains the result of sentimental analysis of movie tweets along with a discussion of the results. At the first step tweets were collected and analyzed from various sentiment analyzers of sentiment analysis with the aim of determining the accuracy of these analyzers. The main purpose of using a different sentiment analyzers such as TextBlob, SentiWordNet, and WSD was to give a comparison of results with different analyzer.

Tweets of movie Race 3 gathered from public accounts were 3000 in number. Among the three sentiment analyzers we compared in this research, we found that TextBlob had the highest rate of tweets with neutral sentiment, 1729 in number and 57.6 % in percentage. SentiWordNet gave 377 & 197 the highest negative & highly negative sentiment rate, 12.5% & 6.6 % respectively, which can be viewed in Table 1 and Figure 3.

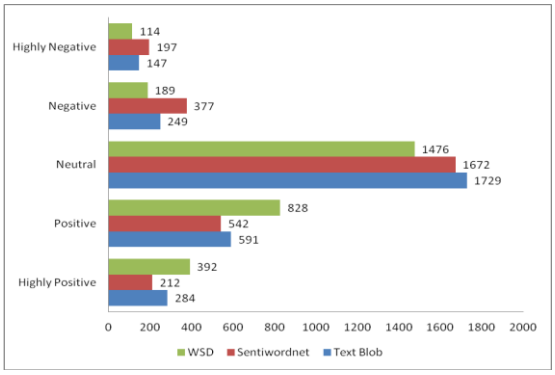


Figure 3. Polarity calculation with each sentiment analyzer.

Table 1. Polarity calculation with sentiment analyzer percentage accuracy.

Sentiment Analyzer	Text Blob	Sentiwordnet	WSD
Highly Positive	284 (9.5 %)	212 (7.07 %)	392 (13.1 %)
Positive	591 (19.7 %)	542 (18.06 %)	828 (27.6 %)
Neutral	1729 (57.6 %)	1672 (55.7 %)	1476 (49.2 %)
Negative	249 (8.3 %)	377 (12.5 %)	189 (6.3 %)
Highly Negative	147 (4.9 %)	197 (6.6 %)	114 3.8 %)

V. CONCLUSIONS AND FUTURE WORK

This paper focuses on the adoption of various sentiment analyzers to determine the approach with the highest accuracy rate for learning about movie opinion. In a lexicon-based sentiment analysis, semantic orientation is of words, phrases or sentences calculated in a document. Polarity in the lexicon-based method is calculated on the basis of the dictionary, that consists of a semantic score of a particular word. Considerable work has been done in the field of sentiment analysis from sentiment lexicons. But, in this research we are providing best adopted sentiment analysis for movie opinion by giving comparison between sentiment lexicons (W-WSD, SentiWordNet, TextBlob). As a result, we calculated sentiments from three analyzers named SentiWordNet, TextBlob, and W-WSD. Although TextBlob results were relatively better as is clearly shown in the results. In order to take our initiative to next level, we will find the patterns of any product or person based on Twitter reviews in future research.

REFERENCES

[1] S. Argamon, K. Bloom, A. Esuli, and F. Sebastiani, "Automatically determining attitude type and force for sentiment analysis," in Human Language Technology. Challenges of the Information Society. Springer, 2009, pp. 218–231.

[2] Boštjan Kaluža, "Machine Learning in Java", first published: Published by Packt Publishing Ltd, UK, 2016.

[3] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, pp. 82–89, 2013.

[4] Kanavos, A.; Nodarakis, N.; Sioutas, S.; Tsakalidis, A.; Tsolis, D.; Tzimas, G. Large scale implementations for twitter sentiment classification. Algorithms 2017

[5] Kanavos, A.; Perikos, I.; Hatzilygeroudis, I.; Tsakalidis, A. Emotional community detection in social networks. Comput. Electr. Eng. 2017, 65, 449–460.

[6] U. R. Hodeghatta, ''Sentiment analysis of Hollywood movies on Twitter,'' in Proc. IEEE/ACM ASONAM, Aug. 2013, pp. 1401 1404.

[7]Twitter Apps. Available online: http://www.tweepy.org/ (accessed on 26 February 2018).

[8] TextBlob Docs. Available online https://textblob.readthedocs.io/en/dev/

[9] Miller, G.A.Wordnet: A lexical database for english. Commun. ACM 1995, 38, 39–41.

[10] Hogenboom, A.; Van Iterson, P.; Heerschop, B.; Frasincar, F.; Kaymak, U. Determining negation scope and strength in sentiment analysis. In Proceedings of the 2011 IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK, USA, 9–12 October 2011; pp. 2589–2594.

[11] Ibrahim, M.; Abdillah, O.; Wicaksono, A.F.; Adriani, M. Buzzer detection and sentiment analysis for predicting presidential election results in a twitter nation. In Proceedings of the 2015 IEEE International Conference on Data

Mining Workshop (ICDMW), Atlantic City, NJ, USA, 14–17 November 2015;pp. 1348–1353.

[12] Roesslein, J. Tweepy Documentation. 2009. Available online: http://docs.tweepy.org/en/v3.5.0/