

Predicting The Public Opinion From The Social Networks

¹Arun.K, ²Dr A. Sri Nagesh, ³P.Praveen

¹Research scholar in CSE department, ACHARYA NAGARJUNA UNIVERSITY, Guntur.

Assistant professor in CSE department, KGR CET, Moinabad, HYD,

²Associate professor, RVR&JC, Guntur,

³Assistant professor in CSE department, KGR CET, Moinabad, HYD,

Abstract: - People in the world connected together and share their opinions, messages, comments, through social networks like facebook, whatsapp, twitter, linked in, Youtube. Mostly these are microbloggings and comments that are from youtube videos which allows limited number of words to post and there is no limit in reading the posts. Social networks have millions of active users per day, to discuss and share their opinions on current news and events that are happening. So this is the best platform to predict the opinions of the public in terms of sentiment analysis. This analysis can give positive and negative opinions of the public, the scaling of the positive (+1 to +5), negative (-1 to -5) and neutral (0) and also emotions. In this paper predicting the sentiment analysis on youtube comments are by using some machine learning techniques and methods in R language. This sentiment analysis can give the decision making for the current events, markets, government policies and products.

Keywords: - Sentiment analysis, NLP, polarity, machine learning, youtube comments, twitter tweets

Introduction

In the 21st century most of the people are using e-social networks to share their messages[10], comments, decisions, likes, dislikes and tweets by using facebook, whatsapp, linkedin, twitter, youtube. It is the best platform to get the public opinion on the topics, comments on the videos, tweets and re-tweets. Now in this model we collect the data from different social networks, pre-process the data as per the requirement then apply it to one of the machine learning methods for getting the polarity of each and every data object. Finally the results will be displayed in the form of graphical representation and these steps that are explained in a sequential approach.

System model

The system model describes the complete flow of data in the sentiment or opinion analysis[8]. The following figure gives the flow of the data in the sentiment analysis for any social network.

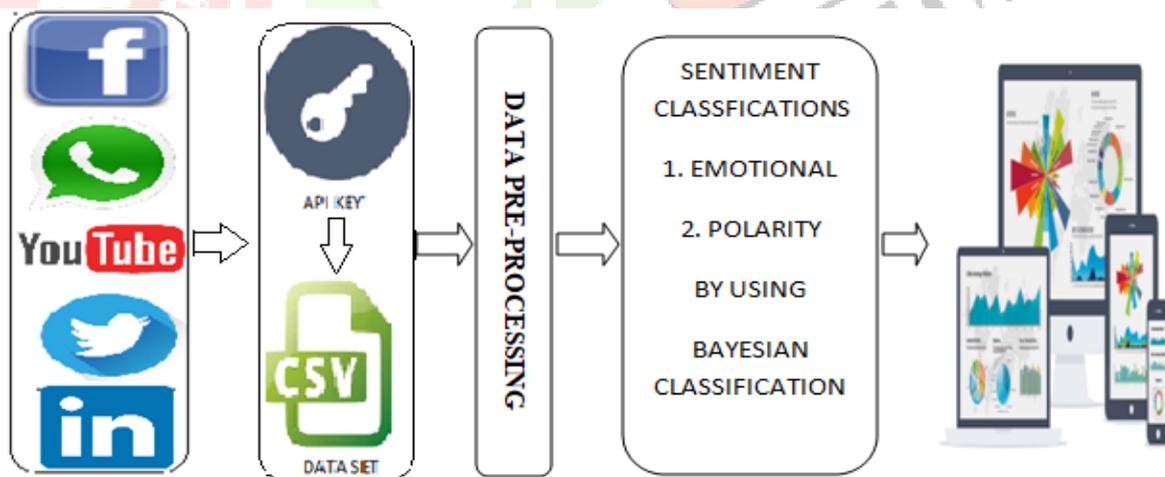


Figure 1: model architecture for sentiment analysis on social networks.

- Social networks** are different apps that are available in the internet or in the smart phones and the information is shared within the groups only. [1]As per the world social networks statistics in the 2017, facebook is in the first place which have the 2.06 billion monthly active users. Youtube is in the second place which has the 1.5 billion active users per month. whatsapp is in the third place which have the 1.3 billion active users per month. Like that twitter is in the ninth place which has the 305 million active users per month and all are used to predict the sentiment analysis.

2. The second phase of the figure is **getting data set through the API key**, the API keys are different for each application. These API keys provides authentication to the users in getting the current data. In R-language sufficient packages are available for accessing the data sets from the facebook, whatApp, twitter[2], youtube[3], linkedin.
3. **Data pre-processing phase** is used to clean data and transform the data into required form of data set [4]. The data set generally contains 1. Web links, 2. Numbers, 3. Tags, 4. Re-tweets, 5. Replies .These are not required for the data analysis, so all the above mentioned should be cleaned along with the unnecessary words in the sentence. Finally this process can hold the required words from each sentence to the classification process. Data pre-processing must conduct different process for different applications. Each application has different tags, links, sequence numbers, replies, re-tweets and common words.
4. **Sentiment classifications** are the heart phase in this structure and there are two types of classifications that are applied in this paper.
 - i) **Emotional sentiment score** is used to represent the sentimental emotions from the pre-processed data set. Emotional sentiments [5] are classified into seven labels like Anger, Fear, Sadness, Surprise, Disgust, and Unknown. Method which is used in emotional sentiment scores:
`classify_emotion(data-set, algorithm="bayes", prior=1.0)`
 - ii) **Polarity sentiment score** is used to classify the data set object into three classes. Such as 1. Positive, 2. Negative, 3. Neutrals. This classification is based on the supervised machine learning approach using naive bayes[7]. The naive bayes classify the given lexemes into numerical scales in between (-5 to +5).
 Positive label: score in between (+1 to +5).
 Negative label: score in between (-5 to -1).
 Neutral label: score is equal to "0".
 Method which is used in polarity sentiment scores:
`Classify_polarity(data-set, algorithm="bayes")`
5. **Graphical Representations:** R-language supports more number of packages for graphical representations as per the requirement we have in consideration.

Results and Experiment:

In this experiment, I consider the sentiment analysis on the youtube comments [6], actually youtube stands second place in the world social networks. I downloaded two youtube comment data sets and applied data pre-processing methods to clean the data sets. Two types of machine learning methods are applied to classify the data in to sentimental classes [9]. One is emotional classification and polarity classifications of the youtube data sets, then these classifications are represented in a graphical manner which is a very best representation. In the following steps procedure is explained in a statistical and graphical manner.

Step 1: Data sets accessed from the youtube.

S.no	Name of the youtube	Links	No of comments
1	National Anthem of India - Jana Gana Mana	https://www.youtube.com/watch?v=Bh26zOjIh9I	106
2	Vandematharam	https://www.youtube.com/watch?v=HZaIG8gtq3Q	112

Table 1: Data sets from youtube.

These data sets are downloaded from the youtube with number of comments posted in those videos.

Step 2: Data sets are applied for the pre-processing and in this process all the unwanted data was deleted.

Step 3: Sentiment analysis: Applied naive bayes for the polarity of the sentiments and the following statistical information is the result.

S.no	Name of the youtube	Positive comments	Negative comments	Neutral comments	Unknown comments	Total no.of comments
1	National Anthem of India - Jana Gana Mana	77	09	11	9	106
2	Vandematharam	103	2	5	2	112

Table 2: Sentiment polarity scores on the youtube Data sets.

Step 4: Represents the emotional sentiments in a graphical manner. In this emotional sentiment there are total seven labels which are used like 1. Joy, 2. Anger, 3. Surprise, 4. Sadness, 5. Fear, 6. Disgust, 7. Unknown. These are represented for one of the data set in the following figure.

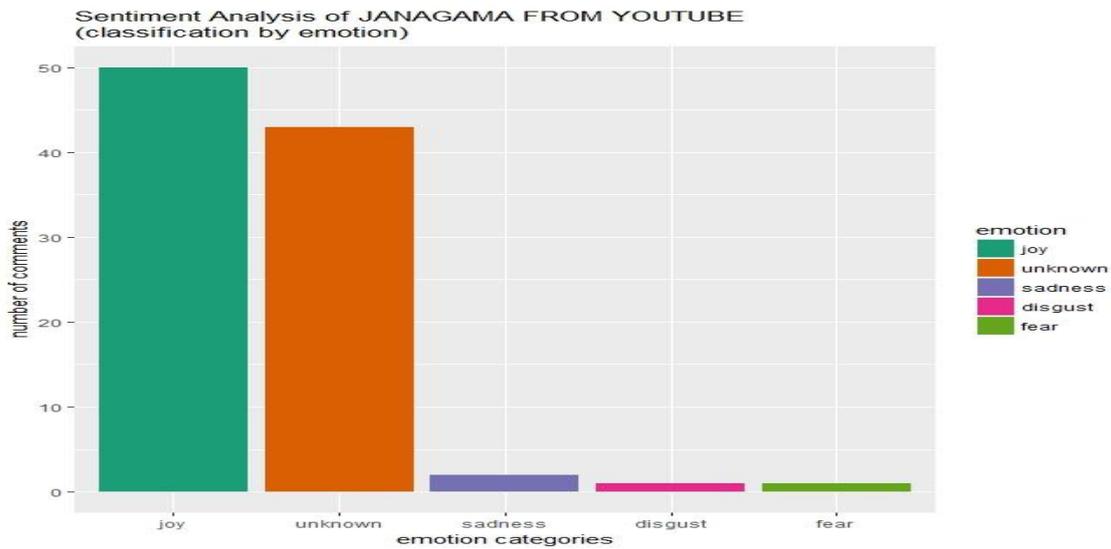


Figure 2: Emotional sentiment scores on the janaganama youtube song.

Step 5: Represents the polarity sentiment output, When ever naive bayes is applied for the pre-processed data. which contains three labels 1. Positive 2. Negative, 3. Neutral. It is for one of the youtube comments which following.

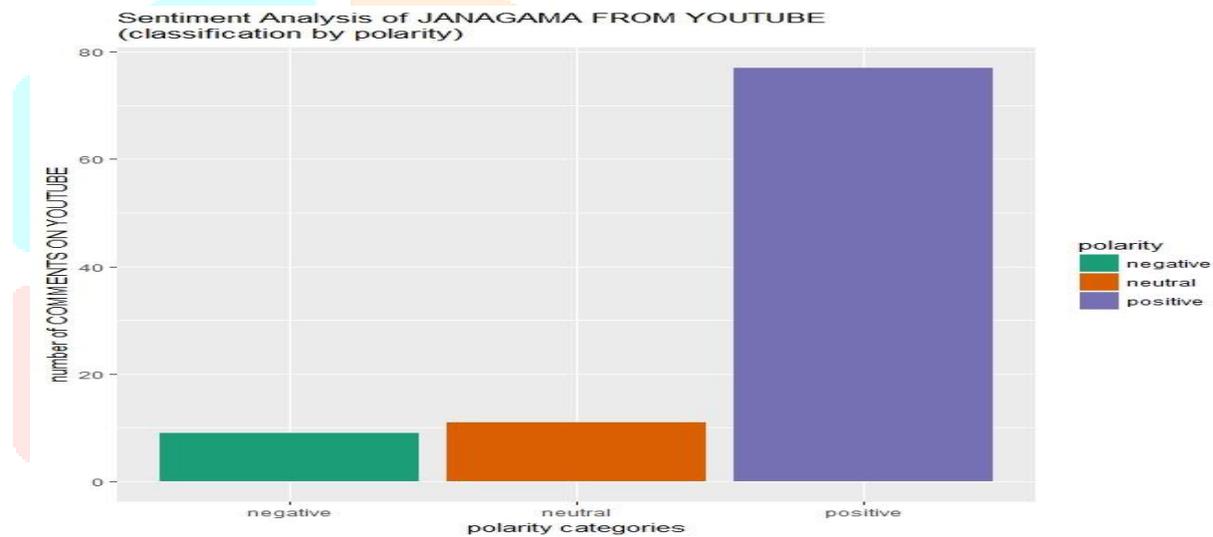


Figure 3: Polarity Sentiment score on Janaganamana youtube song.

Conclusion

In this paper sentiment analysis on the youtube data set is completed with the help of the supervised machine learning method naive bayes. And the same method is applicable to all different social networks to get sentiment analysis, such as twitter, facebook, whatsapp, etc. Unigram, bigrams features are also included in this method. In this way predicting to public opinion is defined. Further it is need to improve the sentiments on the unknown tags in the classified data. Social networks are available in all local language which is very close to the society, so it is better to predict the sentiment analysis on the local languages.

References

- [1] <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>.
- [2] <https://developer.twitter.com/en/docs/basics/authentication/guides/access-tokens>
- [3] https://developers.google.com/youtube/registering_an_application#Create_API_Keys
- [4] Zhao Jianqiang, Gui Xiaolin "Comparison Research on Text Pre-processing Methods on Twitter Sentiment Analysis", 2169-3536 (c) 2016 IEEE, Translations and content mining.
- [5] Kun-Lin Liu, Wu-Jun Li, "Emotion Smoothed Language Models for Twitter sentiment Analysis," AAAI, 2012
- [6] <https://www.curiousgnu.com/youtube-comments-text-analysis>
- [7] Ajay Deshwal, Sudhir Kumar Sharma, "Twitter Sentiment Analysis using Various Classification Algorithms", 978-1-5090-1489-7/16/©2016 IEEE.
- [8] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.
- [9] Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal* 5.4 (2014): 1093-1113.
- [10] Imran, Muhammad, et al. "Processing social media messages in mass emergency: A survey." *ACM Computing Surveys (CSUR)* 47.4 (2015): 67.

