Sentiment Analysis of Twitter Data

Jisha M G
Knowledge Campus, Jain University,
Jayanagar 9th block, Bangalore.

Abstract: The sentiment analysis over twitter is very effective way to observe the feelings of public towards their brands, political news, business, and movies etc. There are several methods to classify the datasets on the blogs. The datasets on twitter is added up with a feature that specifies the positives/negatives on it. The approach of predicting the sentiment has to be applied on the twitter datasets.

Keywords: Sentiment Analysis, Datasets.

INTRODUCTION

Twitter is popular micro blogging site with the maximum 140 characters in length. The tweets are used to express the emotions of tweeter's on any subject. The tweets contain much information about people's opinion. Normally, the tweets are been used by some organizations/ firms to improve their product[2], to get innovation on their field and many other aspects. Sentimental Analysis on twitter data has several challenges since they do not have structured data. The twitter data is streaming form; hence data collected will be accurate. There are two types of data extracted for analyzing the tweets, they are:

1) Data with emoticons which categorizes tweets to positive, negative and neutral based on the tweeter's opinion.
2) The dictionary collected from the web with English translations.

In Twitter, the tweets are posted by the celebrities, politicians of different country, business tycoons, and entrepreneurs. The set of text in twitter can be as below:

1) Positives texts like happiness, enjoy etc. with emoticons like ☺ :D
2) Negative texts like sadness with emoticons ☹ [3]
3) They just right the fact without any emoticons.

Literature survey: Sentimental analysis is handled in many levels, namely document level, sentence level and phrase level. Micro blogging sites like twitter will have real time posts and opinions over it, which has positive and negative reviews. Twitter messages can lead more efficient way of extracting the data. It provides the best way of social communication on an issue, which helps to gather common opinion. The opinion can be classified into polar versus non-polar which means positive versus negative and non-polar4 side can be considered as neutral also.

Data Description: Tweets are the real time messages sent by the micro-blogging users on tweeters - Common place to share uncommon opinions. The terminologies of twitters are:

Tweets: The messages sent by users are tweets.
Ex: Just informed it was National Wine Day...from your insurance agent...take an Uber or Taxi please! #NationalWineDay

Emoticons: Facial expression, represented by the punctuations.
Ex: The Best Song Ever ☺ :*

Target : Users of twitter , use the @ symbol to target on others.
Ex: a pleno sol @Germandicesare @patriciohogan
Hash Tags: used to refer the Main topics.
Ex: Carbonada consistente **#25DeMayo**

Steps:

1) Retrieval of twitter data.
2) Pre-Processing of data.
3) Removing the unwanted Phrase.
4) Analyzing and converting the special characters and emoticons.
5) Getting the final output.

1) *Retrieval of data:*
   The data has to be first taken from the blog to analyze on it.
2) *Pre-Processing of data:*
   The data retrieved had to pre- processed before sending for the extraction, this also includes filtering of data, organizing data to a structured format from its unstructured form.
3) *Removing the unwanted Phrase:* The unwanted data has to be removed (…,%.., *,a,an,the)etc, so that it is easy to refine the data and then convert it during analysis.
4) *Analyzing and converting:*
   The refined data has to be analyzed and converted to the positive, negative and neutral reviews.
   Ex: @, #, ☺ etc
5) *Final output:*
   Once converting of data is done, they can be classified into different classes and deployed.

APPLICATION OF TWITTER DATA:

1) Used to review the related websites, review the movies/ products.
2) To know about the sensitive matters.

3) To know the current trends and fashions.

4) Public opinion on political leaders, business leads, Actors etc.

CONCLUSION:

This analysis helps in improving emoticons and other communication activities. It also increases the quality of the opinion mining and predicting the people's feeling1.Using emoticons, hash tags we collect the useful data for assessment of our analysis.

These symbols will help in easy assumption of the public opinion on the trendiest news or tweets. The analysis on these data helps in better understanding of human mind, how hard they think and what is their opinion on different aspects. The data that are retrieved are classified to positive, negative and neutral classes. We use common process of NLP with will provide us the easy way to derive the context of phrase.

FUTURE WORK:

The purpose of this research includes analyzing the emoticons, symbols and context. But in the future we can even try to analyze the videos, moving emoticons and other complicated contexts also.

REFERENCE:

[1]   http://www.slideshare.net/sumit786raj/sentiment-analysis-of-twitter-data

[2]   http://www.cs.columbia.edu/~julia/papers/Agarwaletal11.pdf

[3]   http://www.aclweb.org/anthology/W12-5504

[4]   https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/download/.../3251