Natural Language Processing

Instructor: Prof. Lu Xiao

Team Members: Removed for Confidentiality

Sentiment Analysis Twitter

Introduction

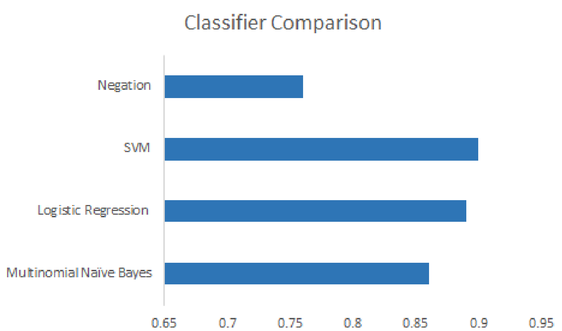
Sentiment Analysis is the process of determining whether a piece of writing is positive, negative or neutral. It' is also known as opinion mining, deriving the opinion or attitude of a speaker. It classifies the polarity of a given text at the document, sentence, or feature/aspect level—whether the expressed opinion in a document, a sentence or an entity feature/aspect is positive, negative, or neutral.

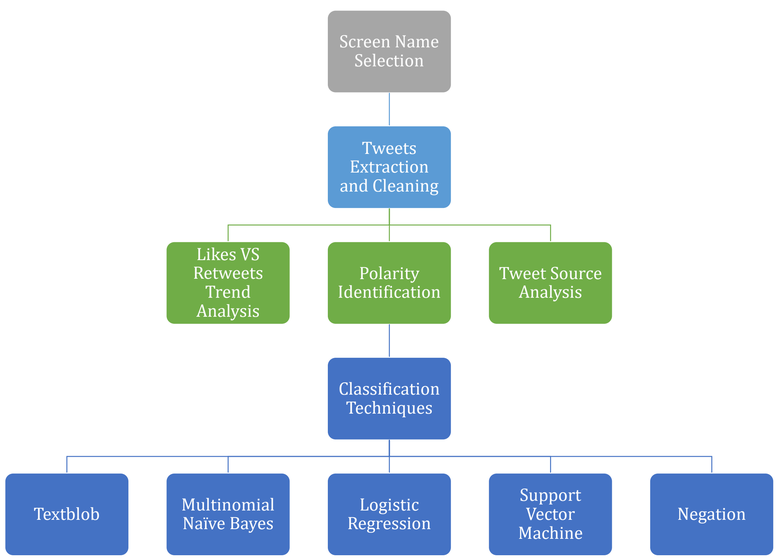
Implemented Techniques

Data Extraction: We have used REST API for our experimental analysis to download a user's tweets, followers, friends and user profile information. We used tweepy library to retrieve tweets from twitter.

Data Preprocessing: Used regular expressions to extract the text from tweets. Stopwords, Special Characters and URL hypertexts were also removed before sentiment analysis.

Data (Sentiment) Analysis: The twitter data was analyzed to study various trends, responses on tweets (likes and retweets) and polarity of sentences. Various NLP techniques were used to classify the tweets into categories. We also implemented language detection for a limited number of tweets. The flowchart represents different Data and NLP techniques used.





Important Definitions

- **Textblob**: It is a Python library used for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of- speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.
- **Featuresets**: Text feature extraction is the process of transforming what is essentially a list of words into a feature set that is usable by a classifier. Here, we extract features from text, to train a classifier.

NLP Classification Techniques:

We found below NLP techniques in our research:

- Naïve Bayes
- Multinomial Naïve Bayes
- Logistic Regression
- Negation classifier using Not Feature set
- Maximum Entropy
- Support Vector Machine
- Tree Kernel Based Model

Suggested Directions

Ideas worth exploring in the future:

- **Affective aspect:** Text which is influenced by or resulting from emotions
- **Lexicon based Sentiment Analysis:** Local context of the word is taken into consideration
- **Subjective Analysis:** Classify parts of text as to whether it was subjective or objective
- **Contextual Analysis:** Analysis of sentiments with respect to the situation and what it means in that situation
- **Fine grained approaches of sentiment analysis:** Emotion, Mood, Attitude, personality traits, subtlety, sarcasm or metaphor

References

[1] Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau. Sentiment Analysis of Twitter Data. Department of Computer Science, Columbia University
[2] Alec Go, Richa Bhayani, Lei Huang. Twitter Sentiment Classification using Distant Supervision. Stanford University
[3] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 79{86, 2002.}