Twitter Sentiment Analysis

Ajay Singh

April 10, 2015

Abstract

I introduce an approach to classify the sentiment behind twitter messages. These messages (obtained via a specific query) are classified into positive or negative. This classification is applicable to tweets about products, individuals, movies or even organizations. The training data has to be very large in order to get the sentiment behind such varied topics.

Motivation

Twitter is a popular microblogging website where users create "tweets", often expressing opinions about different topics. There has been large amount of research in the area of sentiment classification, but that has mostly been limited to reviews on websites, especially about movies and products. Twitter is different because it allows personal feedbacks to be aggregated, that too in a very large number.

Previous Work

Noteworthy work has been done in twitter sentiment analysis by Alec Go, Richa Bhayani and Lei Huang from Stanford University. They used maximum entropy classifier (built from machine learning algorithms), quite different from previous work where keyword-based appraoch was employed. More about their work can be found here.

Approach

The work to be done in this project will be mostly along the lines of the work done earlier, but instead of having a single classifier, I intend to build three classifiers (namely Naive Bayes Classifier, Maximum Entropy Classifier and Support Vector Machines). Also, the dataset used will be large enough (with 1,60,000 tweets distributed equally between positive and negative) to obtain an accuracy as high as possible. The feature extractor will be unigram as using bigrams may make it a little too complex.

References

[1] sentiment140, Go, Bhayani and Huang, Stanford University.

[2] Twitter sentiment classifier using Python and NLTK Laurent Luce.

[3] Naive Bayes Classifier Jacob Perkins.

[4] Twitter Sentiment Niek Sanders