EFFECTIVENESS OF A NAÏVE-BAYES CLASSIFIER FOR SENTIMENTAL ANALYSIS

Dinesh Kannan

kannan.d@husky.neu.edu

PROBLEM DESCRIPTION

Open internet forums hold a lot of text capturing opinion of people across different times. Detecting emotions of these opinions has always been interesting. A common interest has always been to find out if an opinion is 'Positive' or 'Negative'. From a computational perspective, this is a classification task which takes in a piece of text as input and outputs a label, either 'Positive' or 'Negative' denoting the emotion of the text. Sentimental analysis is quite interesting because it provides a simple feedback about what people think. Many businesses use sentimental analysis to gain knowledge about what their customers feel. E.g. Understanding customer reviews on an E-Commerce website. It is also interesting to find out if it is computationally easy or effective to perform such sentimental analysis. This project aims to find out if a Naïve-Bayes classifier (a simple probabilistic classifier) is effective for sentimental analysis. This project also suggests ways to improve the effectiveness of classification tasks for sentimental analysis, by analyzing an improved/modified Naïve-Bayes algorithm or a completely different classification algorithm.

ALGORITHMS

This project trains a Naïve-Bayes classifier with Rotten Tomatoes movie review dataset and tests within the same dataset for accuracy of results. This is a popular dataset contributed by a lot of people. Naïve-Bayes classifier is typically used under the assumption that the attributes are independent. This project is going to use the occurrence of words and word sequences (n-grams).

RESULTS

This project,

- Shows the prediction accuracy of a simple Naïve-Bayes classifier in the test dataset
- Shows the prediction accuracy of an improved version of Naïve-Bayes or another better classifier in the test dataset
- Compares both results and draws conclusions