Assignment 1- INFO284

Syver Storm-Furru, Ådne Forslund and Bjørnar Herland

Performance

The training takes approximately 9-12 seconds for the entire training set. Classifying the test set takes about 20 seconds with single threading. Classifying a single review takes under 0.001 seconds. When training with the train set, classifying the test set has an error rate of 49.61 %, classifying the training set is about the same. However, when training with the test set and testing on the training set, the performance is better, at 43.29 %.

Explanation

When it comes to the speed of the program, multi threading could enhance the run time. The error rate is at some points under 50 %, which is not very reliable. But when training with test set, the performance is some points better. To better the performance for the classifier one could implement the smoothing technique described in the assignment. One could also add a filter that removes neutral words, and also get a better filter for special characters. Another way to improve the classifier could be to weight words by their “positivity”, from classing some reviews as more positive than others, from looking at their score. The positioning of the words in the review could also be used to class the words as more or less positive/negative.