Raw Simulation Results for SVM-Modeled Text Classification

Akshay Agrawal, Shane Leonard

November 14, 2014

Overview

The following tables outline the training and testing results applied to the `medicine_gold` dataset, using classifier `naive_bayes` and the `EdxConfusion` data cleaner. Each section reports the result of processing the raw document in a different way before classifying the processed text as a bag of words.

Original Bag-of-Words Approach

The raw input documents were passed directly to the classifiers with no preprocessing.

Classification results for file ../../Confusion Datasets/medicine_gold ...; using classifier NaiveBayes and data_cleaner EdxConfusion +—————————+————-+————+ | Label | Train: F1 | Test: F1 | +==============+============+===========+ | knowledgeable | 0.737699 | 0.441833 | +————————+—————-+————+ | neutral | 0.928654 | 0.888782 | +————————+—————-+————+ | confused | 0.711252 | 0.367184 | +————————+—————-+————+

Full Classification results for file ../../Confusion Datasets/medicine_gold ...; using classifier NaiveBayes and data_cleaner EdxConfusion +—————————+————-+————+ | Label | Train: F1 | Test: F1 | +==============+============+===========+ | knowledgeable | 0.653983 | 0.487236 | +————————+—————-+————+ | neutral | 0.908834 | 0.879653 | +————————+—————-+————+ | confused | 0.648954 | 0.466741 | +————————+—————-+————+

Full2 Classification results for file ../../Confusion Datasets/medicine_gold ...; using classifier NaiveBayes and data_cleaner EdxConfusion +—————————+————-+————+ | Label | Train: F1 | Test: F1 |

+===============+=============+============+ |
knowledgeable | 0.658809 | 0.484604 | +————————+————————-+————————+ |
neutral | 0.902679 | 0.872805 | +————————+————————-+————————+ | confused |
0.63694 | 0.486843 | +————————+————————-+————————+

Classification results for file ../../Confusion Datasets/medicine_gold . . . ; using classifier NaiveBayes and data_cleaner EdxConfusion +——————————+————————-+————————+ | Label | Train: F1 | Test: F1 | +===============+=============+============+ |
knowledgeable | 0.737699 | 0.441833 | +————————+————————-+————————+ |
neutral | 0.928654 | 0.888782 | +————————+————————-+————————+ | confused |
0.711252 | 0.367184 | +————————+————————-+————————+