

Jordan University of Science and Technology

Performance OF KNN AND SVM Classifiers on Full Word Arabic Article

Authors: I. Hmeidi, B. Hawashin, and E. El-Qawasmeh

Abstract: Abstract This paper reports a comparative study of two machine learning methods on Arabic text categorization. Based on a collection of news articles as a training set, and another set of news articles as a testing set, we evaluated K nearest neighbor (KNN) algorithm, and support vector machines (SVM) algorithm. We used the full word features and considered the tf.idf as the weighting method for feature selection, and CHI statistics as a ranking metric. Experiments showed that both methods were of superior performance on the test corpus while SVM showed a better micro average F1 and prediction time.