

Jordan University of Science and Technology

A Study of the Effects of Preprocessing Strategies on Sentiment Analysis for Arabic Text

Authors: Rehab Duwairi, Mahmoud El-Orfali

Abstract: Sentiment analysis has drawn considerable interest among researchers owing to the realization of its fascinating commercial and business benefits. This paper deals with sentiment analysis in Arabic text from three perspectives. First, several alternatives of text representation were investigated. In particular, the effects of stemming, feature correlation and n-gram models for Arabic text on sentiment analysis were investigated. Second, the behaviour of three classifiers, namely, SVM, Naive Bayes, and K-nearest neighbour classifiers, with sentiment analysis was investigated. Third, the effects of the characteristics of the dataset on sentiment analysis were analysed. To this end, we applied the techniques proposed in this paper to two datasets; one was prepared in-house by the authors and the second one is freely available online. All the experimentation was done using Rapidminer. The results show that our selection of preprocessing strategies on the reviews increases the performance of the classifiers.