Abstract

Sentiment analysis is an area of great interest in research because of its importance and advantages in many different domains. Different approaches and techniques are used to classify the sentiment of texts, and there are different algorithms proposed to improve the performance through text preprocessing. Stemming is one of preprocessing step that is used in many research to enhance the performance of sentiment classification. In this research, we provide new comparative experiments on the impacts and effects of using two of the most commonly used stemmers in the Arabic language; light stemmer and Khoja stemmer in the preprocessing phase of sentiment analysis. We used large Arabic multi-domain datasets that include positive and negative reviews across multiple domains. Five classifiers are used; Naïve Bayes, support vector machines, k-nearest neighbors, decision trees, and logistic regression. The results indicate that Khoja stemmer outperformed that light stemmer Khoja stemmer in terms of precision, recall, f-measure, and accuracy, and it has an advantage in minimizing training time.