Abstract

Restaurant reviews on online sites often provide reviews with scores that do not match the sentiment of the review. A review could have a low score but the review contain positive meanings and a high-score review could have negative meanings. Various attempts to classify sentiment using sentiment analysis have been carried out in many studies. However, sentiment analysis by relying solely on the supervised learning approach gives the results of one class tends to appear more frequently which mean this decreases the performance of the classifier. In this paper, the Improved Naive Bayes approach, Naive Bayes with unigram and bigram features combined with the approach using lexicon is proposed to improve the performance of classifiers. The feature is obtained by extracting the POS TAG pattern that contains words or phrases that express emotions that are relevant to the restaurant review. The Lexicon is built manually by collecting unigram and bigram words and phrases that show relevant emotions expressed in the restaurant review. Classifiers using Improved Naive Bayes show better performance than classifiers using Naive Bayes. Improved Naive Bayes evaluation obtains precision 80%, recall 77%, and F1 76%. While Naive Bayes evaluation obtains precision 68%, recall of 60%, and F1 56%.

Keywords: bigram, improved naive bayes, lexicon, naive bayes, sentiment analysis, unigram