Abstract

Nowadays, people use the Internet to write their opinions on blogs, social networks, and websites. Review sites like IMDb are one of the most frequently visited sites by Internet users that provide comprehensive information about actors, crew, ratings and reviews of films given by others. The reviews and ratings of these films can influence their buying behavior. Machine learning approach is used to solve the problemof ambiguity of opinion by classifying the film review into a sentiment. In this study, 50.000 datasets from IMDb will be used to test the Support Vector Machine classification with Information Gain feature selection to help the performance of the SVM classification. From the test results, obtained the highest accuracy value of 86.1% for unigram and 76,2% for bigram in SVM classification using Information Gainas the feature selection.

Keywords: SVM, sentiment analysis, Information Gain, movie review, n-grams