

ABSTRACT

Film and movies are part of digital lifestyle. A numerous video streaming platforms appear as consequence. Movie review is an alternative for choosing movies which are going to watch next. Unfortunately, the movie review may contain spoiler which is not good for movie buffs. Actually, they only need to know the sentiment information that occurs in movie reviews. Extracting sentiment information in movie review can be done by sentiment analysis. Sentiment analysis, also known as sentiment categorization, is a study that analyses the subjective information on a specific object.

Unfortunately, the bag of word representation used in sentiment analysis has problems in handling high dimensional feature matrix. This problem can be handled by selecting the feature by feature selection. A good feature is the one that has high relevance to the output class. Current work in feature selection for sentiment analysis succeeds in capturing highly relevant feature, but the occurrence of selected features is rare. As a result, sentiment analysis is facing an over-fitting problem.

Information Gain is a common scoring method to select feature. In this study, a feature selection and classification based on Information Gain is proposed. The proposed feature selection, IGDFFS, selects features which satisfy two criteria: (1) high relevance to the output class and (2) high occurrence. The Information Gain score is also used to build a polarity dictionary. The proposed classification scheme, IGC, uses a dictionary to classify the movie review. The experiment result shows that the combination of IGDFFS and IGC, whose accuracy achieves 96%, is more effective than the other methods proposed in the previous work.

Keywords: Sentiment Analysis, Feature Selection, Classification, Information Gain