

Improving sentiment classification accuracy of financial news using n-gram approach and feature weighting methods

ABSTRACT

Sentiment classification of financial news deals with the identification of positive and negative news so that they can be applied in decision support system to perform stock trend predictions. This paper explores several types of feature space as different datasets for sentiment classification of the news article. Experiments are conducted based on n-gram approach (unigram, bigram and the combination of unigram and bigram) used as feature extraction with different feature weighting methods, while, document frequency (DF) is used as feature selection method. We performed experiments to measure the classification accuracy of support vector machine (SVM) with two kernel methods of linear and Radial Basis Function (RBF). Results showed that an efficient feature extraction increased classification accuracy when it is used as a combination of unigram and bigram. Moreover, we also found that DF can be applied as a dimension reduction method to reduce the feature space without loss of accuracy.

Keyword: Sentiment classification; Financial news; Support vector machine; Radial basis function; Linear kernel; Document frequency; TF-IDF