

Sentiment classification of financial news using statistical features

ABSTRACT

Sentiment classification of financial news deals with the identification of positive and negative news so that they can be applied in decision support systems for stock trend predictions. This paper explores several types of feature spaces as different data spaces for sentiment classification of the news article. Experiments are conducted using N-gram models unigram, bigram and the combination of unigram and bigram as feature extraction with traditional feature weighting methods (binary, term frequency (TF), and term frequency-document frequency (TF-IDF)), while document frequency (DF) was used in order to generate feature spaces with different dimensions to evaluate N-gram models and traditional feature weighting methods. We performed some experiments to measure the classification accuracy of support vector machine (SVM) with two kernel methods of Linear and Gaussian radial basis function (RBF). We concluded that feature selection and feature weighting methods can have a substantial role in sentiment classification. Furthermore, the results showed that the proposed work which combined unigram and bigram along with TF-IDF feature weighting method and optimized RBF kernel SVM produced high classification accuracy in financial news classification.

Keyword: Sentiment classification; Financial news; N-gram models; Traditional feature weighting methods; TF-IDF; Linear kernel SVM; RBF kernel SVM; Document frequency