

Effective method for sentiment lexical dictionary enrichment based on Word2Vec for sentiment analysis

ABSTRACT

Recently, many researchers have shown interest in using lexical dictionary for sentiment analysis. The SentiWordNet is the most used sentiment lexical to determine the polarity of texts. However, there are huge number of terms in the corpus vocabulary that are not in the SentiWordNet due to the curse of dimensionality, which will limit the performance of the sentiment analysis. This paper proposed a method to enlarge the size of opinion words by learning the polarity of those non-opinion words in the vocabulary based on the SentiWordNet. The effectiveness of the method is evaluated by using the Internet Movie Review Dataset. The result is promising, showing that the proposed Senti2Vec method can be more effective than the SentiWordNet as the sentiment lexical resource.

Keyword: Sentiment analysis; Word2Vec; Word embeddings; SentiWordNet