Abstract

The implementation of calculation semantic similarity between word is one of task that can be done by Natural Language Processing. The calculation of semantic similarity between word can used to help the machine to understanding of human language(natural language). Beside that, calculation of semantic similarity can be used as a basic of the next step in NLP's research. The main idea of this study is motivated by a problem where nowaday the seaarch of information sistem are involved by many text and document, so we need to help the machine to understand those texts or documents. A pair of word are similar if they have similarity to the level of meaning of concept. In this research, we are implement the calculation of semantic similarity between word in English. The corpus that used in this research are Brown Corpus, Berita Corpus, and Harry Potter Corpus. That documents are convert into vector space by using Word2vec. Next, the score of semantic similarity generated by vector are compared to SimLex999 Gold Standard dataset to measure their corelation. The result showed that Word2vec have corelation's score of 0.192 in Pearson corelation.

Keywords: Semantic Similarity, Natural Language Processing, Word2vec