

Abstract

Semantic word similarity is a measurement of the extent to which two words have the same meaning. The meaning of the word is different depending of the vocabulary used. Word embedding is a branch of science that studies the human language with a representation of the word into the form to the vector so this vector represents the meaning of those words. In this final task implemented semantic word similarity with one of the methods of word embedding that is word2vec using data sets English language and Spanish language of corpus. Research results form this final task is the value of the correlation between the value of semantic word similarty by the system with the gold standard. The highest correlation value is 0.238 using size 200 for the same window size. The parameters that affect the value of semantic word similarity with word2vec method is the size of the vector.

Keywords: Word2Vec, gold standard, correlation
