Abstract

Semantic relatedness refers to the degree to which the two concepts or words associated (or not), while the semantic similarity is a special case or subset of a semantic relatedness. The similarity of words (word similarity) is a measurement of how closely a pair of words semantically, if a word pair have synonyms relationship then they has the highest value. Pointwise Mutual Information (PMI) is a statistical measurement of the semantic relatedness and semantic similarity that has been widely used. One variant of the PMI is Second Order Co-ocurrence pointwise Mutual Information (SOC-PMI). The results on this research is the correlation between similarity scores generated by the gold standard system SimLex-999, WordSim353 and Miller and Charles. The highest correlation value is 0.2881 by using window size = 33 and a value of δ = 6.5. The parameters that cause the best correlation with PMI-SOC method is term-context between words that compared.

Keywords: Semantic Similarity, Pointwise Mutual Information, Second Order Co-ocurrence pointwise Mutual Information