

Abstract

The similarity between two sentences is the score that represents the relatedness and likelihood between those sentences. Measuring sentence similarity has attracted researchers over the last few years, although determining the performance and effectiveness of measuring sentence similarity is a key issue. The similarity between sentences can be syntactic or semantic; syntactic similarity is based on string matching, while semantic similarity implies similar meaning between sentences. Sentence representation in the distributional space involves measuring sentence similarity based on the similarity between vectors. In this section, we propose a new approach that measures sentence similarity by including lexical, semantic, and syntactic–semantic features and applying linear regression and support vector machine regression. To evaluate the performance of the proposed approach, experiments are conducted on two datasets, namely, Arabic paraphrasing benchmark, MSRvid, and SMTeuroparl from SemEval 2017. The proposed approach achieves a correlation of 0.354 when applied to Arabic paraphrasing benchmark because of the complexity of this dataset, while it achieves a correlation of 0.743 and 0.467 on the MSRvid and SMTeuroparl dataset, respectively.