

Adapting gloss vector semantic relatedness measure for semantic similarity estimation: an evaluation in the biomedical domain

ABSTRACT

Automatic methods of ontology alignment are essential for establishing interoperability across web services. These methods are needed to measure semantic similarity between two ontologies' entities to discover reliable correspondences. While existing similarity measures suffer from some difficulties, semantic relatedness measures tend to yield better results; even though they are not completely appropriate for the 'equivalence' relationship (e.g. 'blood' and 'bleeding' related but not similar). We attempt to adapt Gloss Vector relatedness measure for similarity estimation. Generally, Gloss Vector uses angles between entities' gloss vectors for relatedness calculation. After employing Pearson's chi-squared test for statistical elimination of insignificant features to optimize entities' gloss vectors, by considering concepts' taxonomy, we enrich them for better similarity measurement. Discussed measures get evaluated in the biomedical domain using MeSH, MEDLINE and dataset of 301 concept pairs. We conclude Adapted Gloss Vector similarity results are more correlated with human judgment of similarity compared to other measures.

Keyword: Semantic web; Similarity measure; Relatedness measure; UMLS; MEDLINE; MeSH; Bioinformatics; Pearson's Chi-squared test; Text mining; Ontology alignment; Computational linguistics