Semantic Similarity and Relatedness Evaluation Based on Ontological Meta-Properties

Alcides Lopes[1]

[1]Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, Brazil

Abstract

The human's capability of comparing entities in the real world is outstanding. Although the natural cognitive ability applies semantic similarity or relatedness measures for most decision-making processes straightforwardly, reproducing this ability in computer systems is still a research topic. Also, these measures are widely applied in information retrieval, natural language processing, machine learning, etc. In the last decades, researchers in Computer Science have proposed theories, approaches, methods, and metrics for semantic similarity and relatedness evaluation. This evaluation uses a knowledge resource (e.g., dictionaries, taxonomies, ontologies, etc.) from which a semantic measure extracts its similarity or relatedness assumptions. For example, the path-based semantic measures assume that the number of relationships between two entities expresses the degree of relatedness between them. However, the state-of-the-art semantic measures ignore the advances in the ontology modeling field. One of these advances is using top-level ontologies or ontological meta-properties to model the ontology entities according to a well-founded category system. In our view, by ignoring this advance in ontology modeling, we are led to two problems: the state-of-the-art semantic measures can be negatively affected by the structure of these ontologies; the state-of-the-art semantic measures are not developed to support the used top-level ontology or the ontological meta-properties during similarity or relatedness evaluation. Thus, we use these two previous problems as the motivation of this doctorate work. In this work, we aim to establish a methodology for using the ontological meta-properties of top-level ontology concepts to improve semantic similarity and relatedness evaluation between domain entities. In our view, we can establish links between the discussion of ontological meta-properties and the theory of similarity and relatedness. For example, we can match the theory about the semantic similarity with the discussion about the identity between entities where two entities share the same identity if they are the same. In addition, we aim to propose a framework to support the use of our methodology in information retrieval and natural language processing systems. So far, we have started research on improving the time performance of the similarity evaluation between terms that represent ontology entities. Since a term can represent many ontology entities in a given text, state-of-the-art algorithms for term

FOIS 2021 Early Career Symposium (ECS), held at FOIS 2021 - 12th International Conference on Formal Ontology in Information Systems, September 13-17, 2021, Bolzano, Italy

agljunior@inf.ufrgs.br (A. Lopes)

0000-0003-0622-6847 (A. Lopes)

similarity evaluation need to measure similarity between all these senses and select the one that reached the highest similarity value. In order to reduce the number of comparisons, we only evaluate the similarity between two terms in which their respective ontology entities specialize top-level concepts that are not disjoint. We evaluate this methodology on three gold-standard datasets for term similarity: Miller&Charles, Rubenstein&Goodenough, and MEN, with 28, 65, and 3000 term pairs, respectively. In our experiments, we achieve a reduction from 10.8 seconds to 3.8 seconds, from 25.8 seconds to 7.5 seconds, and from 42.7 minutes to 7.1 minutes, respectively, without losing the correlation with human judgments of the traditional algorithm for similarity evaluation.

Acknowledgements

The author gratefully acknowledge the financial support of the Brazil Federal Agencies CAPES and CNPQ, and the grant and scientific cooperation of PETROBRAS company.