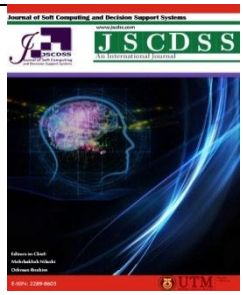
An International Journal
JSCDSS
http://www.jscdss.com
Vol.2 No.6 December 2015: 1-13
Article history:
Accepted 14 September 2015
Published online 14 September 2015

Journal of Soft Computing and Decision Support Systems



E-ISSN: 2289-8603

A Review of Semantic Similarity Measures in Biomedical Domain Using SNOMED-CT

Mojtaba Zare [a,*], Christina Pahl [a], Mehrbakhsh Nilashi [a], Naomie Salim [a], Othman Ibrahim [a]
[a] Faculty of Computing, Universiti Technologi Malaysia, Johor, Malaysia

*** Corresponding author email address**: mojtabazare123@yahoo.com

Abstract

The determination of semantic similarity between word pairs is an important task in text understanding that supports the processing, classification and structuring of textual resources. In the field of biomedical, semantic similarity measures have been the focus of much research by exploiting knowledge sources such as domain ontologies. SNOMED-CT as a main biomedical ontology provides a global and broad hierarchical terminology for clinical data storage, encoding, and the retrieval of health and diseases information. In this study, we classified the measures proposed in biomedical domain and used SNOMED-CT as an input ontology. We also examined the studies that evaluated these methods using biomedical benchmarks. Regarding this, three major databases, including Science Direct, Springer and IEEE were selected to extract studies which proposed similarity measures and used SNOMED-CT as a knowledge source. The purpose of this study is to provide the reader with the understanding about the application of semantic similarity measures in biomedical domain using SNOMED-CT, and to gain a clear insight about the performance of these methods. This study also supports researchers and practitioners in effectively adapting semantic similarity measures in SNOMED-CT and provides an insight into its state-of-the-art.

Keywords: Biomedical ontologies, SNOMED-CT, Semantic similarity measure

1. Introduction

Semantic Similarity Measures (SSMs) estimate the similarity between two given concepts (Janowicz et al., 2015; Liao et al., 2014; Sahni et al., 2014). The estimation of the semantic similarity between concepts helps in better understanding of textual resources (Song et al., 2014, Chakraborty et al., 2014). These measures are mainly categorized into two groups, including distributional based and knowledge based methods (Garla and Brandt, 2012a). Distributional based methods utilize the distribution of concepts within a corpus in conjunction with a knowledge source to compute similarity; these measures include corpus Information Content (IC) and context vector methods (Jiang et al., 2014). Knowledge based methods, on the other hand, utilize knowledge sources, such as ontologies and semantic networks. Knowledge based methods are divided into two groups, path-based and intrinsic IC-based measures (McInnes and Pedersen, 2015; Harispe et al., 2014).

Semantic similarity measures have been used in wide array of applications in biomedical domain, using biomedical ontologies. They have been applied to design information retrieval algorithms (Chaves-González and MartíNez-Gil, 2013; Uddin et al., 2013), to disambiguate texts (McInnes and Pedersen, 2013; Miller et al., 2012), to suggest drug repositioning (Gottlieb et al., 2011; Lamurias et al., 2013) and to cluster genes, according to their molecular function (Pesquita et al., 2009; Guzzi et al., 2012). Semantic similarity measures are indeed critical components of many knowledge-based systems (Chang and Lee, 2015; Gottlieb et al., 2011). In addition, they are nowadays receiving more attention due to the growing adoption of both Semantic Web and Linked Data paradigms (Bizer et al., 2009; Iosif and Potamianos 2015).

Semantic similarity measures, based on knowledge sources and ontologies, use the taxonomical evidences modeled in the ontology to assess the similarity of two given concepts. In fact, ontologies support these measures to model unstructured and heterogeneous information through the hierarchical vocabularies and structured sets of concepts (Harispe et al., 2014; Cross et al., 2013; Meng et al., 2013). Fortunately, the field of biomedical has been very prolific in creating medical ontologies which organize concepts in a non-ambiguous way to be used by semantic measures (Batet et al., 2011; SáNchez and Batet, 2011; Al-Mubaid and Nguyen 2009). Some well-known examples of ontologies in biomedical domain include Medical Subject