Advances in Intelligent Systems and Computing, Volume 1073, 2020, Pages 407-419

A Semantic Taxonomy for Weighting Assumptions to Reduce Feature Selection from Social Media and Forum Posts

Hasan, A.M., Rassem, T.H., Noor, N.M., Hasan, A.M.

Faculty of Computing (FKOM), University Malaysia Pahang, Gambang, Kuantan, DarulMakmur, Pahang 26300, Malaysia

Abstract

Numerous researchers have worked on the knowledge-based semantics of words to clarify the ambiguity of (https://github.com/alimuttaleb/Ali-Muttaleb/blob/master/Synonym.txt) synonyms in various natural-language processing fields, such as Wikipedia, websites, and social networks. This paper attempts to clarify ambiguities in the lexical semantics of taxonomy in social media. It proposes a new knowledge-based semantic representation approach that can handle ambiguity and high dimensionality issues in text mining. The proposed approach consists of two main components, namely, a feature-based method for incorporating the relationships between lexical sources and a topic-based reduction method to overcome high dimensionality issues. These components help weight and reduce the relevant features of a concept. The proposed approach captures further lexical semantic similarity between words. It also evaluates the use of (https://wordnet.princeton.edu) WordNet 3.1 in text clustering and constant weighting assumption in the feature-based method used to select concepts/words from social media. To address ambiguity, the semantics of concepts with small feature subset size reduction are represented, and the performance of the semantic similarity measurement is improved. The proposed method evaluates word semantic similarity using the (https://github.com/alimuttaleb/semantictaxonomy/blob/master/mc30.txt) MC30 dataset in WordNet and obtains the following results for semantic representation: r = 0.82, p = 0.81, m = 0.81, and nz = 0.96.

Keywords

Semantic taxonomy; Feature-based method; Semantic representation; Feature selection; Gloss; Social media; MC30