Lecture Notes in Electrical Engineering, Volume 621, 2020, Pages 471-483

A Proposed Method Using the Semantic Similarity of WordNet 3.1 to Handle the Ambiguity to Apply in Social Media Text

Hasan, A.M.[a], Mohd Noor, N.[a], Rassem, T.H.[a], Mohd Noah, S.A.[b], Hasan, A.M.[a]

[a] Faculty of Computing, University Malaysia Pahang, Gambang, Kuantan, Pahang Darul Makmur 26300, Malaysia
[b] Faculty of Information Science and Technology, University Kebangsaan Malaysia (UKM), Bangi Selangor, 43600, Malaysia

Abstract

The semantic similarity between two concepts is widely used in natural language processing. In this article, we propose a method using WordNet 3.1 to determine the similarity based on feature combinations. This work focuses on overcoming the ambiguity in social media text via the selection of informative features to improve semantic representation. In addition, this research uses social media as its research domain used in this work, and the study is only limited to the politic dataset. A feature-based method is applied to predict the outcome and improve the performance of the proposed method depending on factors related to the fidelity, continuity, and balance of knowledge sources in WordNet 3.1. Semantic similarity measurements among words are insufficient and unbalanced features. However, this study presents a semantic similarity measure of a feature-based method in WordNet 3.1 to determine the similarity between two concepts/words depending on the selected features used to measure their similarity, which is also known as a "noun" and "is-a" relations-based method. We evaluate our proposed method using the data set in Agirre [1] (AG203) and compare our results of our new method as which three of methods taxonomy relation, non-taxonomy and Glosses with those of related studies. The correlation with human judgments is subjective and low based on our results was a better. Experimental results show that our new method significantly outperforms other existing computational methods with the following results: *r = 0.73%, p = 0.69%, m = 0.71% and nonzero = 0.95%*.

Keywords

Semantic similarity; Semantic relatedness; Taxonomy relations; Glosses; Semantic representation; WordNet 3.1; Social media; Proposed method; AG203