



Knowledge-Based Semantic Relatedness measure using Semantic features

Ali Muttaleb Hasan¹, Noorhuzaimi Mohd Noor^{2*}, Taha. H. Rassem³, Ahmed Muttaleb Hasan⁴

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

¹alimatlab65@yahoo.com, ^{2*}nhuzaimi@ump.edu.my, ³tahahusseini@ump.edu.my, ⁴ahmed.matlab11@gmail.com

ABSTRACT

Measuring semantic relatedness has received much attention for uses in many fields such as information retrieval and natural language processing. For handling synonymous problem in distributional-based measures, many researchers are investigating how to exploit semantic features in lexical sources to form knowledge-based measures. In the knowledge-based measures, a hierarchy model is used to measure the relatedness between words based on only the taxonomical features extracted from a provided lexical source. In this paper, a new knowledge feature-based measure is proposed to build the semantic vector of a word construct on taxonomical and non-taxonomical feature of relation words. The proposed measure utilised the topological parameters that weight the importance of each element in the semantic vector. One of the gold dataset used to assess the proposed model and compare the findings with other related works. The results demonstrated the effectiveness of the proposed model on measuring semantic relatedness between words. In this paper, the research framework is identified based on the observations made on the previous related works that have been conducted for semantic representation and semantic relatedness measures. The required data in this research includes the semantic knowledge-based approach and the evaluation datasets. The semantic knowledge that will be used throughout of this research is extracted from English WordNet 3.1. On the other hand, the evaluation datasets covers the gold standard benchmarks which have been used for evaluating the semantic relatedness measurements and text mining tasks. Finally, the evaluation is preform to evaluate the proposed method (PM) based on approach in this research, in which obtained the result have been analyzed, to discuss and compare based on different performance measure and finding the strength and weakness in this paper, to alternative the semantic representation correlated to this research, to designing and develop the topical-based on the semantic representation method for text mining from Social media.

Keywords: semantic representation, semantic similarity, semantic measure, topological parameter, lexical source,

1. INTRODUCTION

The semantic similarity is a metric of defined sets documents or terms of words, several metrics are used ¹WordNet 3.1, as a manually construct lexical of the source of English words. Despite the advantage of having human supervision in construction of the database, since the words are automatically to learn the database, so that is not the measure of relatedness between multi-words. The semantic relationship between units of language such as concepts or instances. The words of semantic similarity are often confused with semantic relatedness. The semantic relatedness includes any relation between two words of concept, while the semantic similarity only includes "is a" relations. E.g., "car" is similar to "bus", but is also related to "road" and "driving". Computationally, semantic similarity can be estimated by defining a topological similarity, by using ontologies to define the distance between words/concepts. Few investigations have used feature strategies to measure the semantic relatedness of words' meanings. The main goal of feature-based techniques in deciding semantic similitude depends on the features chosen to represent the semantics of concepts and the type of measurements used to measure the likeness between two delegate features. The result of few studies [1] [2] [3] [4] shows that the semantic similarity heavily relies on the features of a selected concept. In lexical sources, semantic features can be placed into two main categories: graph-based and feature-based [5].

In the graph-based method, a concept's meaning is represented as either semantic taxonomy or semantic ontology. The main idea behind this method comes from cognitive science the human brain depends on linking concepts to form the semantics of a given concept [6] [7]. When the brain receives a concept, it recalls other related concepts and links these concepts together to understand the degree of relationship between the meaning of the received concept and other concepts. A proposed method and

This work is supported by the University of Malaysia Pahang (UMP) via research grant UMP RDU1803141. Corresponding author: Noorhuzaimi Mohd Noor^{2*}, nhuzaimi@ump.edu.my^{2*}. machine learning, NLP.

¹ <https://wordnet.princeton.edu>