

A Word-to-Vector Formulation for Word Representation

Authors : Sandra Rizkallah, Amir F. Atiya

Abstract : This work presents a novel word to vector representation that is based on embedding the words into a sphere, whereby the dot product of the corresponding vectors represents the similarity between any two words. Embedding the vectors into a sphere enabled us to take into consideration the antonymity between words, not only the synonymity, because of the suitability to handle the polarity nature of words. For example, a word and its antonym can be represented as a vector and its negative. Moreover, we have managed to extract an adequate vocabulary. The obtained results show that the proposed approach can capture the essence of the language, and can be generalized to estimate a correct similarity of any new pair of words.

Keywords : natural language processing, word to vector, text similarity, text mining

Conference Title : ICPRL 2018 : International Conference on Pattern Recognition and Machine Learning

Conference Location : Madrid, Spain

Conference Dates : March 26-27, 2018