عنوان مقاله:

Benefiting from Structured Resources to Present a Computationally Efficient Word Embedding Method

محل انتشار:

مجله هوش مصنوعی و داده کاوی, دوره 10, شماره 4 (سال: 1401)

تعداد صفحات اصل مقاله: 11

نویسنده:

.F. Jafarinejad - *Faculty of Computer Engineering, Shahrood University of Technology, Shahrood, Iran*

خلاصه مقاله:

In recent years, new word embedding methods have clearly improved the accuracy of NLP tasks. A review of the progress of these methods shows that the complexity of these models and the number of their training parameters grows increasingly. Therefore, there is a need for methodological innovation for presenting new word embedding methodologies. Most current word embedding methods use a large corpus of unstructured data to train the semantic vectors of words. This paper addresses the basic idea of utilizing from structure of structured data to introduce embedding vectors. Therefore, the need for high processing power, large amount of processing memory, and long processing time will be met using structures and conceptual knowledge lies in them. For this purpose, a new embedding vector, WordᵧNode is proposed. It uses a well-known structured resource, the WordNet, as a training corpus and hypothesis that graphic structure of the WordNet includes valuable linguistic knowledge that can be considered and not ignored to provide cost-effective and small sized embedding vectors. The NodeᵧVec graph embedding method allows us to benefit from this powerful linguistic resource. Evaluation of this idea in two tasks of word similarity and text classification has shown that this method perform the same or better in comparison to the word embedding method embedded in it (WordᵧVec). This result is achieved while the required training data is reduced by about ۵۰,۰۰۰,۰۰۰%. These results provide a view of capacity of the structured data to improve the quality of existing embedding methods and the resulting vectors.

کلمات کلیدی:

Word Embeddings, WordNet, Word Similarity, Graph Embeddings, NodeᵧVec

لینک ثابت مقاله در پایگاه سیویلیکا:

https://civilica.com/doc/1570953