Paper: *Evaluation methods for unsupervised word embeddings*

Razwan Ahmed Tanvir

Date: 01/25/2022

Quote *"We present a novel evaluation framework based on direct comparisons between embeddings that provides more fine-grained analysis and supports simple, crowd-sourced relevance judgments."*

Overview There are several word embedding models that provide meaningful vector representation of words and these vector representations have linguistic relationship between them. However, there exists no direct evaluation method or framework of these models. This paper in discussion, from the authors of Cornell, presented a complete study on evaluation methods for unsupervised embedding techniques.They surveyed the evaluation methods of the embedding models not the embedding methods itself. They categorized existing evaluation methods in two divisions, namely, Extrinsic and Intrinsic evaluation. However, the authors identified some drawbacks in these methods and consequently proposed their framework that address the existing problems.

This research generate some promising results indicating that word embeddings models encode information about word frequency. Furthermore, it sheds doubt on the mostly used cosine similarity for similarity measure of the vector space. The author released a frequency-calibrated query inventory with the judgment of all the users.

Intellectual Merit This paper carries out extensive study to provide a framework for the evaluation of the unsupervised embedding techniques, and it expands the existing knowledge of the word embedding models and their evaluation. This paper formulates a reference point for these embedding models. This research is notable for the natural language processing based applications for its applicability. This paper presents results for six embedding models on 14 different datasets. They showed their results on four different categories- Relatedness, Cetegorization, Selectional Preference, and Analogy tasks. The results of this research are well reasoned and showed with justification and the method of this research is well organized and transparent. The authors of these papers are from Cornell and they are well capable of running this research activities. The authors use several datasets to train their models which include- November 2007 snapshot of Wikipedia, Wikipedia dump (2008-03-01). They mentioned that they used Stanford tokenizer to tokenize their data.

Broader Impact This research produces framework for word embedding models which boosts the performance and reliability of language processing applications. The evaluation of word embeddings will guide further research to test their model against this framework which will in turn bring progress in natural language processing tasks in general. This research is referenced in other research works a lot which says a lot about the impact of this research. The authors are former Cornell academics. They released several resources including code and datasets. However, the authors of this paper are all male.

Keywords Natural Language Processing, Word Embedding, Model Evaluation

Discussion Questions

- In the discussion part the authors mentioned that their results mean that cosine-similarity for the intrinsic tasks get polluted by the frequency based effects, and to me is not clear and the author did not present any explanation on that. I think there should be specific claim about how the embedding space is affected by the word frequency.
- Authors claimed that there further researches need to be done to unleash the linguistic relationship between words and I think that we should put more focus on establishing linguistic relationship in natural language in general, apart from the words' vector representations.

Table 1: Grade deductions by section

Overview	Intellectual M.	B. Impact	Keywords	Questions	Is Online?