Paper: *Efficient Estimation of Word Representations in Vector Space*

Mushfika Rahman

Date: 21/1/2022

Quote *... there are situations where simple scaling up of the basic techniques will not result in any significant progress, and we have to focus on more advanced techniques.* pp.12.

Overview The authors of this paper solved the problem of learning high-quality word vectors from massive datasets while tackling computational complexity. Before the word2vec model, previously proposed models had been trained on a few hundred million words, and they were successful with modest dimensions. They proposed two new models derived from the techniques used by the previous model, the NNLM model. One architecture is the Continuous Bag-of-Words model, and another is the Continuous Skip-gram Model. The architecture in both models is named log-linear, which saves computation significantly. The CBOW model predicts target words from the neighboring context while Skip-Gram implements the opposite task. The Skip-Gram model predicts neighboring context from target words(2). The proposed models by the authors were able to capture similarities between words in terms of syntactic and semantic similarities. It maintains linear regularities like differences in syntax and semantics, letting to implement computing analogies such as vector addition and cosine similarity.

The authors suggested techniques for measuring the quality of the resulting vector representations that will project similar words that tend to be close to each other and that words can have multiple degrees of similarity (1). They also implemented Parallel Training of Neural Networks that significantly reduced training time on the extensive dataset and resulted in high accuracy.

Intellectual Merit The researchers at Google came up with the word2vec model. Before their proposed model, in NLP, treated words as atomic units did not capture the relationship between words. The existing models had a significant limitation that the performance is dominated by the size of high-quality transcribed speech (1). The proposed models were derivative of the existing model, where the model is trained on top of distributed representation of words. The researchers also proposed a methodology to determine the quality of the resulting vector. The proposed methodology takes advantage of complex machine learning models. The research advanced in the field of Natural Language Processing.

Broader Impact The research can be successfully applied to Knowledge Discovery and verify the correctness of existing facts (1). It impacted the various fields in NLP-based systems such as machine translation, chatbot, and many more. Many of the work has been published after the word2vec model. There is another implicit impact on the recommendation system on non-textual data, which can help the business grow. The researchers later published the multi-thread code for computing vectors in both architectures. They also published more than 1.4 million vectors that represent named entities, trained on more than 100 billion words(1).

Keywords Computing methodologies, Natural language processing, Machine learning, Unsupervised learning, Neural Network

Discussion Questions

- One of the unexpected findings during this experiment was the resulting vectors can be used to answer very subtle semantic relationships between words(1). Such as it can answer the question. For instance, what is the word that is similar to *'big'* in the same sense as *'small'* is similar to *'smaller'*? It could be answered by simple algebraic operation. (1)
- One of the potential limitations would be the semantic prediction of a word relies only on its surrounding context. It may sometimes lead to sub optimal solutions.

References

[1] Efficient Estimation of Word Representations in Vector Space,T Mikolov, K Chen, G Corrado, J Dean

[2] https://israelg99.github.io/2017-03-23-Word2Vec-Explained