

Title

Efficient Estimation of Word Representations in Vector Space

1st Author

Tomas Mikolov - Tomas Mikolov

Submission Date

16 Jan 2013

Publisher

-

You can download this paper from [here](#)

Abstract:

we propose two novel model architectures for computing continuous vector representations of words from very large data sets. The quality of these representations is measured in a word similarity task, and the results are compared to the previously best performing techniques based on different types of neural networks. We observe large improvements in accuracy at much lower computational cost, i.e., it takes less than a day to learn high quality word vectors from a 1.6 billion words data set. Furthermore, we show that these vectors provide state-of-the-art performance on our test set for measuring syntactic and semantic word similarities.

My View:

Previously, words were considered as atomic units and the relationship between them was just the similarity; two words are similar or not. However, this paper improves the relationship between words and goes beyond similarity such as synonym. By building up the vector representation of words, words can have multiple degrees of similarity. By some methods, like increasing the size of data set and developing the models, it observed large improvements in accuracy with lower computational cost.

Previous Idea:

- Feedforward neural network with linear projection layer and a non-linear hidden layer.
- where the word vectors are first learned using neural network with a single hidden layer. The word vectors are then used to train the NNLM.

Previous Idea weakness:

Previous models like LDA and LSA had some issues, for example LDA is computationally very expensive or new architectures outperform than LSA for preserving linear regularities.

Paper solution:

The authors aim to maximize the accuracy, while minimizing the computational cost by some methods. For instance, they propose *Parallel Training of Neural Networks* or using *hierarchical softmax* in the output layer. In addition, one of their solutions was *New Log-linear Models* which their training process is divided into two steps to decrease the cost of non-linear layer.

Result:

This paper described the quality of vector representations of words derived by various models on a collection of syntactic and semantic language tasks. It achieved to train high quality word vectors and more accuracy compared to popular neural network models.

Honestly, nothing cross my mind :)