

Paper: *GloVe: Global Vectors for Word Representation*

Mushfika Rahman

Date: 24/1/2022

Quote	... the two classes of methods are not dramatically different at a fundamental level since they both probe the underlying co-occurrence statistics of the corpus, but the efficiency with which the count-based methods capture global statistics can be advantageous. pp.12.
Overview	<p>Methods to learn word embedding before <i>GloVe</i> are successful in capturing the semantic and syntactic meanings in words using vector operation, but how these meanings are emergence remained vague. The most popular word2vec depends only on local language information, which sometimes may lead to sub-optimal solutions. The authors in this paper are trying to overcome these problems by using a global log-bilinear regression model, which takes advantage of two existing modeling approaches that are <i>Global Matrix Factorization</i> and <i>Local Context Window</i>. The method is created on the idea of deriving semantic relationships between words from the co-occurrence matrix. They further derive a cost function that minimizes the difference between the dot product of the vectors of two words and the logarithm of their number of co-occurrences.(1)</p> <p>The model generates a vector with meaningful substructures that performed 75% in the word analogy task. It also produced improved results on word similarity and named entity recognition.</p>
Intellectual Merit	The researchers at <i>Stanford Artificial Intelligence Laboratory and the Natural Language Processing Group</i> introduced the global log-bilinear regression model. The existing Global Matrix Factorization Methods performed poorly on the word analogy task. On the other hand, other models such as CBOW skip-gram had inefficient usage of statistics. They used the transformative concept, which efficiently leveraged statistical information by training only on the nonzero elements in a word-word co-occurrence matrix. The model was evaluated over already existing datasets, including the English Benchmark dataset for NER <i>CoNLL-2003</i> . Their experiment included an extensive dataset with included around 400,000 most frequent words. They further used ten negative samples. Their improved result impacts various Natural Language Processing applications such as information retrieval, document classification, named entity recognition.
Broader Impact	The GloVe model still serves as a benchmark in various NLP areas. It also generates the vectors faster, which is beneficial for model training purposes using a single thread. It is also mentioned that the model could be easily parallelized across multiple machines. The research involved postdoc and students from <i>Stanford University</i>
Keywords	Computing methodologies, Natural language processing, Machine learning, Unsupervised learning, Neural Network
Discussion Questions	<ul style="list-style-type: none">• One of the surprising findings in the model implementation is that while performing semantic sub-task, the model performs better when trained on smaller corpa.

- One interesting analogy is that when we manipulate all the training hyper-parameters, the embeddings produced by the word2vec and GloVe tend to perform similarly or not.

Table 1: Grade deductions by section

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

References

- [1] <https://paperswithcode.com/method/glove>
- [2] GloVe: Global Vectors for Word Representation, Jeffrey Pennington and Richard Socher and Christopher D. Manning