

# GloVe: Global Vectors for Word Representation

## 1 Review

Point of view [1] about this paper, One of the strengths of this paper is the thoroughness of the evaluation section, where the authors compare the performance of GloVe to other state-of-the-art word embedding methods on a variety of tasks. This helps to establish the effectiveness of GloVe and its suitability for a wide range of NLP applications. Additionally, the authors provide a detailed description of the algorithm itself, including the various hyperparameters that can be tuned to achieve optimal performance. This level of detail makes it easy for other researchers to replicate the experiments and build upon the work presented in the paper. One potential limitation of this paper is that the evaluation is largely focused on word-level tasks, and it may not generalize as well to larger-scale tasks such as document classification or machine translation. However, this is a common limitation of word embedding methods in general, and it is not unique to GloVe.

From [2] The two main model families for learning word vectors are: 1) global matrix factorization methods, such as latent semantic analysis (LSA) is efficiently leverage statistical information, and poorly on the word analogy task, indicating a sub-optimal vector space structure. 2) local context window methods, such as the skip-gram model of Mikolov may do better on the analogy task, but poorly utilize the statistics of the corpus since train on separate local context windows instead of on global co-occurrence counts. In this paper author analyze the model properties necessary to produce linear directions of meaning and argue that global log-bilinear regression models are appropriate for doing so and propose a specific weighted least squares model that trains on global word-word co-occurrence counts and thus makes efficient use of statistics as evidenced by its state-of-the-art performance of 75percent accuracy on the word analogy dataset.

Author introduced a new model for word representation name GloVe, for Global Vectors, because the global corpus statistics are captured directly by the model. After considering so many terms a new weighted least squares regression model introduce a cost function:

$$J = \sum_{i,j=1}^V f(X_{ij}) \left( w_i^T \hat{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \quad (1)$$

and more general weighting function.

Author trained SG<sup>†</sup> and CBOW<sup>†</sup> using the word2vec tool. The text for details and a description of the SVD models when size 1.6B then model GloVe semantic accuracy 67.5 percent and syntactic accuracy 54.3 percent and total 60.3 percent. When the size is 42B then 81.9 percent, 69.3 percent and 75.0 percent respectively. As their primary focus on analogy task, they also include WordSim-353, MC, RG, SCWS and RW for variety of word similarity tasks. For named entity recognition they trained models on CoNLL-03 training data on test on three datasets: 1) ConLL-03 testing data, 2) ACE Phase 2 (2001-02) and ACE-2003 data, and 3) MUC7 Formal Run test set. Total of 437,905 discrete features were generated for the CoNLL2003 training dataset with 50-dimensional vectors for each word of a five-word context are added and used as continuous features.

For the five different word similarity datasets the accuracy of GloVe score WS353- 75.9, MC- 83.6, RG- 82.9, SCWS- 59.6, RW- 47.8 and the size is 42B. F1 score on NER task with 50d vectors model HPCA test score- 88.7, model GloVe score Dev-93.2, Test-88.3, ACE-82.9, MUC7-82.2 and conclude that the GloVe vectors are useful in downstream NLP tasks. A context window that extends to the left and right of a target word will be called symmetric, and one which extends only to the left will be called asymmetric. For showing performance on the word analogy task for 300-dimensional vectors trained on different corpora and on the syntactic subtask, there is a monotonic increase in performance as the corpus size increases. This is to be expected since larger corpora typically produce better statistics.

The run-time is split between populating X and training the model. Populating X with a 10 word symmetric context window, a 400,000 word vocabulary, and a 6 billion token corpus takes about 85 minutes. For

300-dimensional vectors with the above settings (and using all 32 cores of the above machine), a single iteration takes 14 minutes. Overall accuracy on the word analogy task as a function of training time, which is governed by the number of iterations for GloVe and by the number of negative samples for CBOW and skip-gram. All cases 300-dimensional vectors on the same 6B token corpus with the same 400,000 word vocabulary, and use a symmetric context window of size 10. For the same corpus, vocabulary, window size, and training time, GloVe consistently outperforms word2vec. GloVe, is a new global log-bilinear regression model for the unsupervised learning of word representations that outperforms other models on word analogy, word similarity, and named entity recognition tasks.

## References

- [1] OpenAI, “Chatgpt,” 2022.
- [2] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” 2014.