

GloVe: Global Vectors for Word Representation by Pennington, Jeffrey et al.

GloVe was developed at Stanford's NLP Group in 2014. It is a systematic, unsupervised approach for learning word representations in vector space. According to the paper, it outperforms word analogy, word similarity, and named entity recognition tasks by training across aggregated word-word co-occurrence within a corpus of text. GloVe captures meaning in vector space and does so by addressing **global** information as a whole instead of paying attention only to the local bits. Because it trains across co-occurrence, it is designed to predict the co-occurrence ratio between words. Unlike traditional models that train by optimizing loss functions using gradient descent, GloVe weighs loss on word frequency.

To illustrate an example, take the phrase "He ate the apple." Traditional models would locally assign a connection between "the" and "apple", without knowing if "the" is a stopword or has a meaningful connection to "apple". Via GloVe, the vector space captures the dimension of meaning by assigning greater loss to common words such as "the". In order for this to work, a co-occurrence matrix is first built between words, and "meaning" is given to the co-occurrence ratio between two words in a context. The issue presented is the lack of weight associated to co-occurrence pairs, as they are all weighted equally. This presents issues when co-occurrence does not appear as frequently and thus assigns lower meaning scores. Word2vec, an existing related model they discuss in the paper that positions words into vector space on account of their common contexts, was outperformed by GloVe in both speed and accuracy of results using iterations of GloVe vs. negative samples of the continuous bag-of-words and skip-gram models of word2vec.