

Language is an essential tool which has helped humans to achieve prowess over other animals. Hence, it is vital for enabling machines to understand language so that we can move towards general artificial intelligence. One of the most critical concepts in the linguistics is the denotational semantics, which says that a word is determined by the terms surrounding it. This is what we will be using to train our neural network.

Since computer only understands numbers, we represent every word as a one-hot encoding vector. We try to learn dense vector weights for each word.

For training, <https://arxiv.org/pdf/1301.3781.pdf> we try to predict the surrounding words given the centre word. For that, we build the likelihood function and try to minimize the log-likelihood function. In this way, finally, we get our weight vectors. One important thing to understand here is that this is a sort of "false learning" because we are training the model on a different task than the task; we finally want to achieve.

There are several improvements over the vanilla Wrd2Vec, commonly used is the "negative sampling" approach. In this approach, we ignore some of the very frequently words like "the", "is", "a" because they seldom impart any meaning to the neighbouring terms. This makes training less costly and efficient.

References:

1. CS224n Lecture 1
2. <http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/>
3. <https://arxiv.org/pdf/1301.3781.pdf>