**11.05.2020**

Working of Vocabulary Creation for the paper :Dependency Based Word Embeddings

Source code link : <https://code.google.com/archive/p/word2vec/> (mentioned in paper)

-Load the file containing text

-It reads all words and keeps track of their frequency.

-Infrequent words are removed from the vocabulary.

-To search if word is already in vocabulary or not they use hash.

-The vocabulary consists of words and their respective counts(sample vocabulary generated by their code uploaded).

-Create a binary huffman tree for obtaining binary encodings for each word.

-Perform k means cluster for training.

But for dependency contexts we might need something like :

Example: The clock is green

Then an input might be (green nsubj clock).

We should be ignore some dependecies like root, punc,compund.

Since wiki dependency and dependency generated by stanford nlp different in their textformat. We first might need to have both of them common and then work on loading data.

**18.05.2020**

For training in addition to measuring similarity of the object with their cosine product , there could be an extra value in output vector which could define the stability radius.

We propagate this radius measure only when we encounter a preposition in the sentence.

For eg, cup on the table. We measure their similarity with their corresponding embedding using already given method but at the same time have a value which will be denoting their radius as an extra value added to their respective vectors.

The skip gram model gives the output as the probability of the word that is going to closest to given word. So we need to add or modify the model to consider an extra set of weights for just learning this stability value. This set of weights will only be learning whenever there is a preposition encountered. The objective would be to maximise the difference between two values and that the stabler object has a higher value than the less stabler part so like in above example, we have <table> have bigger radius than <cup>.