

Aditya Yele
111447727
NLP Assignment 1

Implementation details

1. Batch Generation

For batch generation I am taking I create two arrays batch and label of the dimension of batch_size.

Find the middle index which will be the context word using by data_index(this is the current index start point of the window)+ skip_windows

Then gather samples by iterating over the range from middle-numskips/2 to middle+numskips/2.

Add all the words in this range except the context word to the labels and corresponding batches array is filled with the middle word

Move the window forward by one and the next word becomes the context word

2. Loss Implementation

a. Cross Entropy

The vector representation of the context words in the inputs array are multiplied with the representations of the neighbouring words. Take only the diagonal elements of this as they are the multiplications corresponding to $v_c \cdot u_o$

Take exponential of this product and log of it this will be A.

Then take the exponential of the dot product of labels and the transpose of inputs Sum it across columns and take the log. Adjust dimensions this will be B.

Subtract A from B

b. NCE

First find all the vector representations of the word indexes in label and sample along with their probabilities and biases.

Then multiply the input and label vectors and take the diagonal elements and add the biases.

Multiply the positive probabilities by the number of samples and add by epsilon to prevent the na values after using log. Subtract this from the diagonal elements of the product of inputs and labels.

Find the sigmoid of the above and take its log. This is A

Multiply the probabilities of the negative words with their number and take its log after adding a small epsilon

Subtract these above probabilities from the multiplication of the negative samples with the context words with added biases

Take the sigmoid of the above vector subtract 1 from it and take its log

Sum it across rows. This is B

Add A and B and multiply with -1.

3. Word analogy

Read the dev sample file line by line .

Parse each line and from the dictionary get the index and find its representations in the embeddings using the index.

For each of the pairs before the || subtract their vectors and find the average of the difference.

For each of the pairs after the || subtract their difference and find the cosine similarity between the vectors of differences and the average difference of the vectors of the words before the ||.

Then find the pair with the maximum cosine similarity it will be the one which is the most related to the given input and find the pair with the least cosine similarity, it will be the one with least relationship.

4. Compute top 20 words

For each of the words. American, would and first iterate through all the words in the dictionary and find the cosine similarity between them.

Print the words which have the maximum cosine similarity with each of them.

Experiment details

The cross entropy model produced by below configuration

200001	128	16	2	0.001	cross	64	33.9
--------	-----	----	---	-------	-------	----	------

The nce model produced by below configuration

400001	128	4	8	0.001	cnce	64	33.9
--------	-----	---	---	-------	------	----	------

NCE model on an average performs better by giving a better accuracy on most configurations.