Assignment 3 Report

Yaseen

March 2024

1 Hyper Parameters

1.1 SVD

for SVD, context windows 1,3,5 were experimented with varying amount of data from top 20,000 rows of the train.csv file to all and the rows of csv file,also different embedding dimensions were experimented with like 128,512.

1.2 Word2Vec

for Word2Vec, context windows 1,3,5 were experimented with only top 20,000 rows as the training took huge amounts of time. Two embedding layers were used one for word embedding, other for context embedding , and the loss function used was nn.CosineEmbeddingLoss() with label being +1 for +ve words, and -1 for -ve words, negative words were sampled using the following distribution,

$$P(w) = \frac{count(w)^{0.75}}{\sum count(w_i)^{0.75}}$$

twice the positive words are selected for negative sampling. the optimizer is Adam Optimizer with learning rate of 0.01 for 5 epochs.

1.3 DownStream Task

For DownStream Task , LSTM was used and nn.CrossEntropy() loss function was used with Adam optimizer with learning rate of 0.001 run for 10 to 20 epochs. for LSTM hidden dimension is 512 dimension with output dimension of 4.

1.4 Results

A general trend in svd, is increasing the no of rows taken for generating the co-occurence matrix increased the accuracy. However in skipgram increasing the context window helped increase accuracy as it directly number of samples both +ve and -ve as -ve samples are sampled twice as much.

Vectors	context window	embedding_dimension	rows	accuracy
SVD	1	128	20,000	55.38%
SVD	3	128	20,000	61.23%
SVD	5	128	20,000	64.75%
SVD	1	512	50,000	63.38%
SVD	3	512	50,000	70.37%
SVD	5	512	50,000	75.76%
SVD	1	512	1,19,000	75.23%

Table 1: SVD results

Vectors	context window	embedding_dimension	rows	accuracy
skipgram	1	128	20,000	70.38%
skipgram	3	128	20,000	73.27%
skipgram	5	128	20,000	78.34%

Table 2: skipgram results