Title: Neural Language Model
Module: COM6513 Natural Language Processing
Name: Ziyuan XIA
Registration Number: 180216466

Description:

In this lab, a neural language model is built. The model uses word embedding in which includes 16 words (including starting mark '<s>' and ending mark '.'). The script is edited and compiled under Windows OS.

The built neural network contains three layers:

1. Input layer with the size of context size times embedding size. The embedding size is set to 2 which is the size of the bag of word to the power of 0.25. The context size is set to 2. In this case, the input layer length is 4.

$$\textit{Input Layer} \sim (2, 2)$$
$$\textit{Input Layer} \sim (1, 4) \text{ (flatten)}$$

2. Hidden layer with the dimension of 128. The Linear 1 function multiplies a (4, 128) matrix with the input layer (1, 4) which makes the size of hidden layer a (1, 128) vector.

$$\textit{Hidden Layer} \sim (1, 4) \cdot (4, 128) = (1, 128)$$

3. Linear function 2 multiplies a (128, 16) matrix with hidden layer 1 (1,128) which makes the output layer a (1, 16) vector.

$$\textit{Output Layer} \sim (1, 128) \cdot (128, 16) = (1, 16)$$

The output layer contains the log probability of each word. The predicted target can be retrieved by:

$$\textit{argmax (Output Layer)}.$$

Result and Discussion:

The sanity check returns success in 5 consecutive runs when the learning rate is 0.03 and the epoch is 85. The result can be reproduced by activating the comment at line 48 in the script (lab5.py). When the sentence is completed successfully, the script prints a heart shaped mark.

The model only predicts 'mathematician' when the context is ['<s>', 'The']. This is because the word 'mathematician' goes after ['<s>', 'The'] occurs so many times that

makes ‘mathematician’ the dominating choice.

Furthermore, every trigram from the training set has been checked so that the accuracy of the model prediction can be obtained. The accuracy changing graph regarding to training epoch is shown in figure 1.

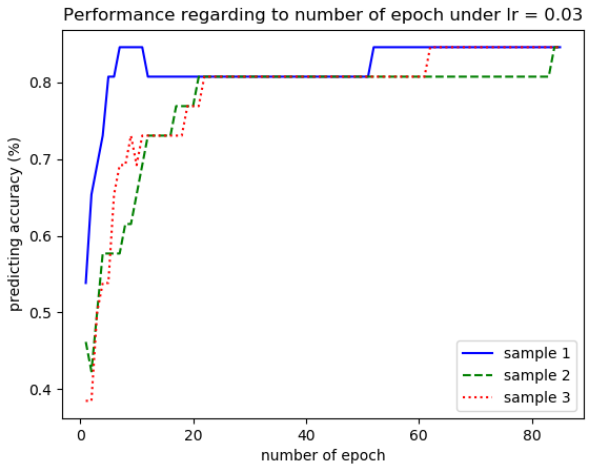

Fig.1 Accuracy changing

The accuracy reaches and keeps at the peak after 85 epochs under lr = 0.03. Only when the accuracy is maximized, the model can 100% success in sanity check. The reason why accuracy cannot reach 100% is that there are dominating choices in this model (like the word ‘mathematician’). One way to avoid this is having a larger training set.

The model can predict “The _____ solved the open problem.” correctly by adding up the log probability of all trigram from the sentence “The philosopher solved the open problem.” and “The physicist solved the open problem.” and choose the one with higher log probability. The model can predict correctly when the accuracy reaches peak in which case the learning rate is 0.03 and number of epochs is 85. The correct answer is ‘physicist’, the embedding of ‘physicist’ is closer with ‘mathematician’ than that of ‘philosopher’ does. This can be proved by the cosine similarity, the cosine similarity between ‘physicist’ and ‘mathematician’ is 0.96 while is -0.47 between ‘philosopher’ and ‘mathematician’ (can be reproduced by using a random seed).

The bigram ML model cannot answer this question because it only considers one context. The model cannot make prediction between unseen choices while neural language model has this ability because the embedding gives a log probability for every target. For an easier understanding, neural network model can be considered make every target an object and every context an attribute which makes it ‘understands’ the meaning of natural language.