DL for NLP - Ass. 2 - 1

Itamar Trainin - 315425967

December 2019

1 Best model Parameters

NER

Embedding layer: #_of_unique_words x 50
Hidden later: 50 x 300
Output layer: 300 x #_of_unique_label

Learing rate: 0.01
Batch size: 32
Iterations: 30

Window size: 5

freq_bound: 2
balance_coef: 1.3

POS

Embedding layer: #_of_unique_words x 50
Hidden later: 50 x 50
Output layer = 50 x #_of_unique_label

Learing rate: 0.1
Batch size: 512
Iterations: 150

Window size: 5

freq_bound: 2
balance_coef: 1.3

2 Considerations

a.

In order to deal with the case where some words appears only in the dev set, and therefore will not be learned in the training, I have replaced 100 least frequent words with '-unknown-' thereby learning an embedding for this new word. Now, if an unseen word was tackled upon, it is replaced with the embedding of the '-unknown-' word.

b.

In order to have a window for all words in the text, I have padded the whole text with the word '-pad-' for NER and '´´' for POS and gave it a label (which is really irrelevant since they never appear as a center word in a window).

3 Graphs

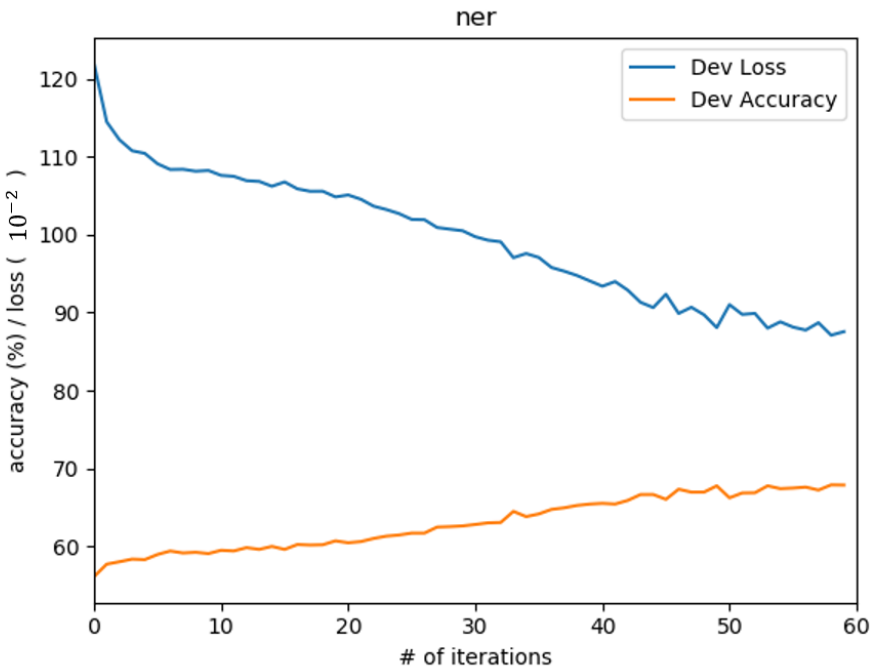

Figure 1: NER

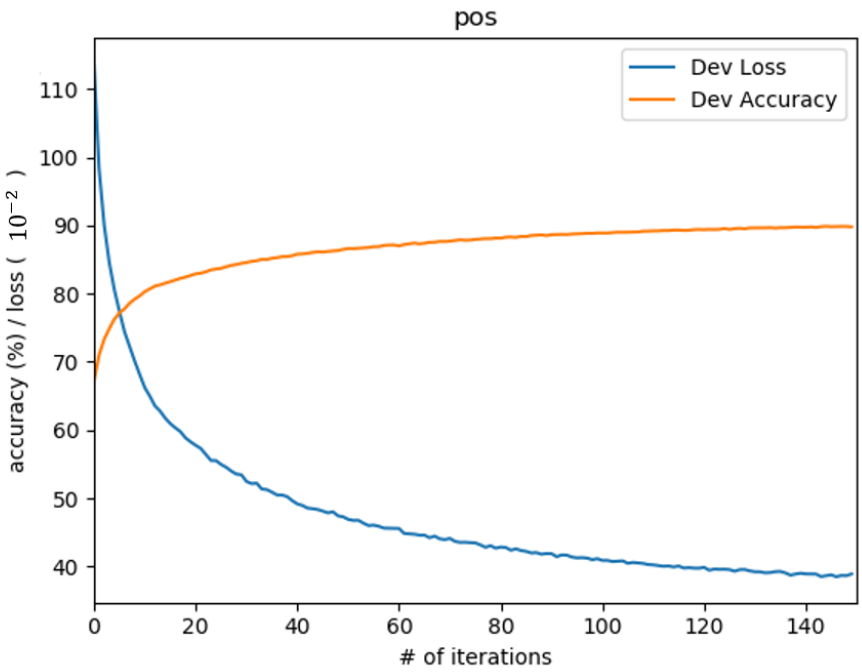
pos
Dev Loss
Dev Accuracy
accuracy (%) / loss (10^{-2})
110
100
90
80
70
60
50
40
0
20
40
60
80
100
120
140
# of iterations

Figure 2: POS