

Part 3

Name

Simon Raviv

Considerations

Best Parameters

NER Tagging

- Epochs: 22
- Hidden layer size: 64
- Batch size: 128
- Learning rate: 0.001
- Learning rate scheduling gamma: 1 (no scheduling in this case)
- Learning rate scheduling step size: 10 (no scheduling in this case)
- Dropout probability: 0.5
- Weight decay: 0.0001

POS Tagging

- Epochs: 45
- Hidden layer size: 80
- Batch size: 128
- Learning rate: 0.009
- Learning rate scheduling gamma: 0.8
- Learning rate scheduling step size: 10
- Dropout probability: 0.2
- Weight decay: 0.00001

Accuracy:

In general, there wasn't big difference in the accuracy at the end. There was a difference in the training process. Using the pretrained embedding, the accuracy was higher for the several first epochs. Which make senses, since the pre trained embedding vectors have been already trained on some other task and are better than random initialized.

NER Tagging:

The dev accuracy (without considering common tag O) is ~81.3 [%].

POS Tagging:

The dev accuracy is ~94.1 [%].

Pre-trained Logic

Words Not in the Vocabulary

The embedding logic uses the embedding vectors for the words appear in the pretrained embedding vectors. For words that are not in the embedding vectors, add embedding vectors initialized to with the normal distribution.

Non-Capital letters Words

I have experimented with 2 approaches:

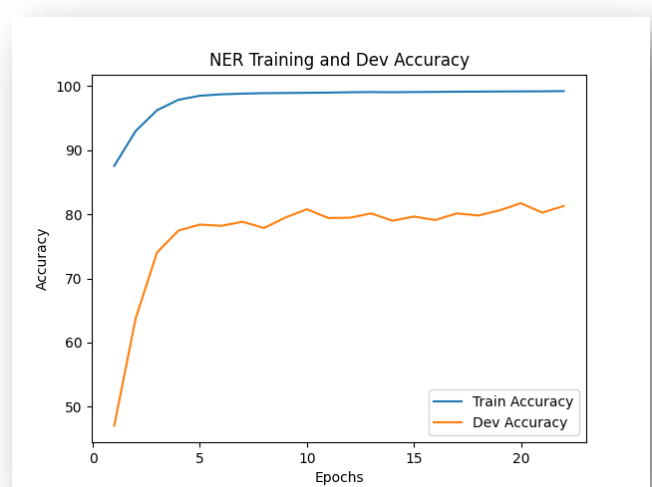
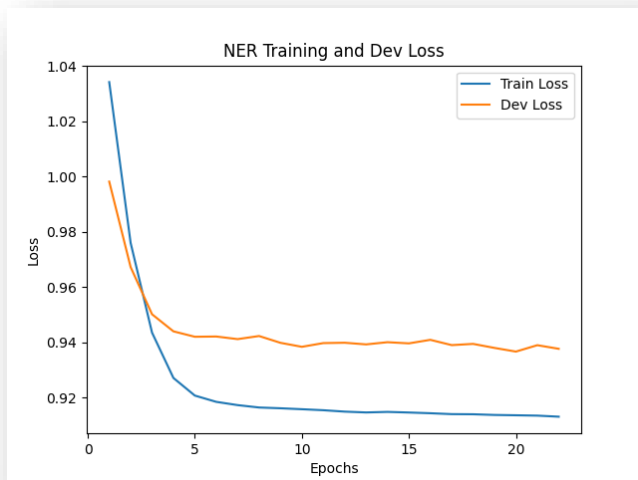
1. Lower case the training set vocabulary.
2. Add new embedding vectors for the words with capital letter in the training set, using the same preceding logic.

I have chosen the second approach, because it showed better results, though not by much. A possible explanation might be that capital letter do matter, for example word in the beginning of a sentence will start with capital word, it can help with understanding its part of speech, names and locations will start with capital letters, so it can also help with recognizing the entities, etc.

Training Graphs

NER Tagging

Note: The calculation for the dev done without considering the common label 'O' as requested. The calculation for the train done with it.



POS Tagging

