Jonathan Mandl 211399175

Danielle Hodaya Shrem 208150433

Part 3 - Adding the external word embeddings

Architecture

In this section, we used the same hyperparameter configuration to both our POS and NER tasks, since it delivered strong performance on each. Our model is a simple feed-forward neural network with a single hidden layer of 256 neurons. The training hyperparameters were:

- Learning rate: 0.001
- Epochs: 5
- Batch size: 64

Considerations

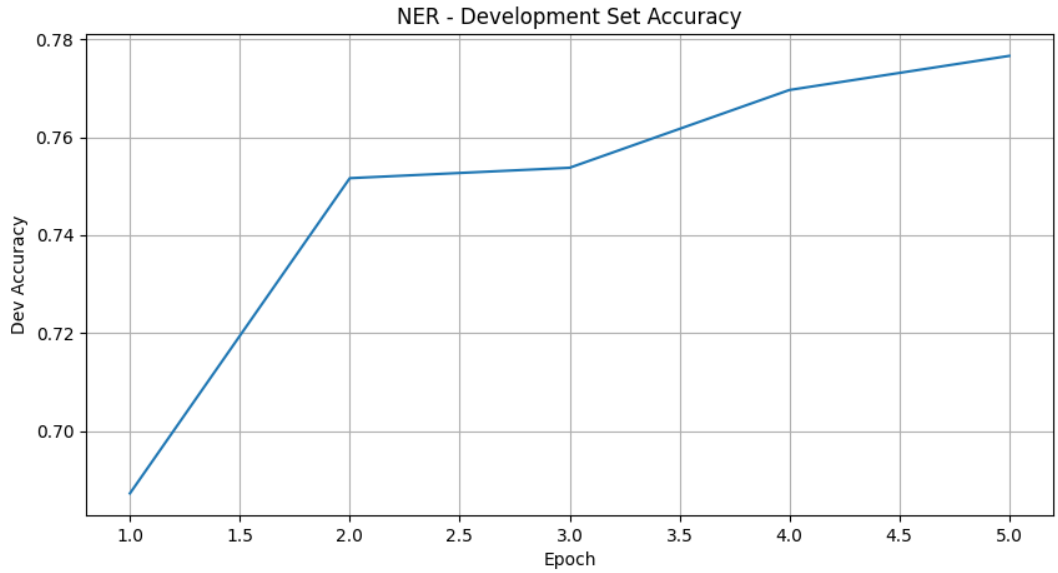
To utilize the pretrained word vectors, we first lowercased every token in our training vocabulary—ensuring it matched the lowercase vocabulary of the pretrained embeddings. We then extracted from the pretrained embedding matrix only rows corresponding to words in our (now lowercase) training set to initialize our embedding matrix. Any training set word that lacked a pretrained vector was assigned a random embedding.

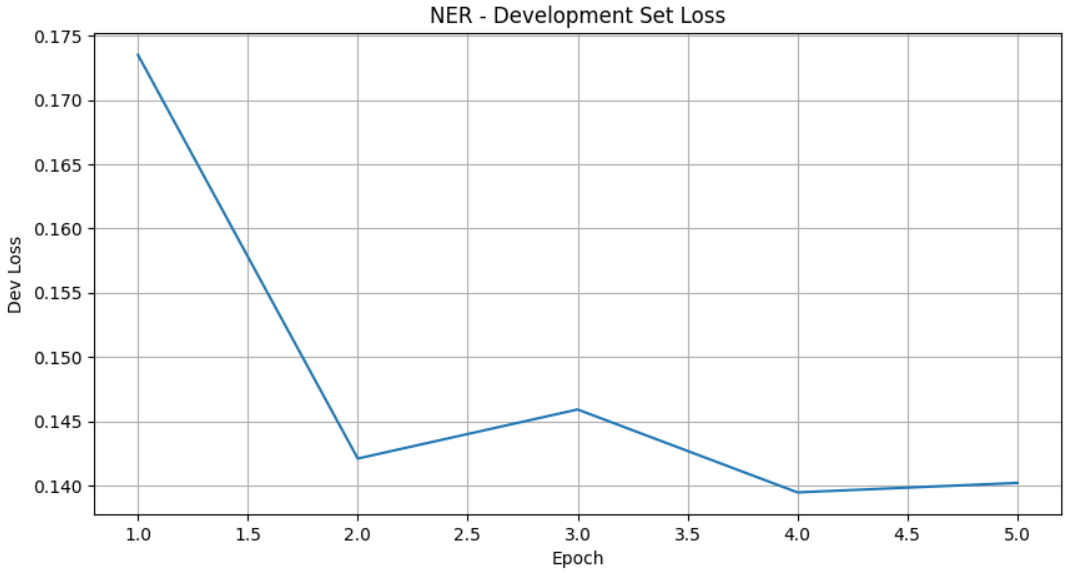
We handled words that appeared in the training set but not in the development set using the same UNK token training as in Part 1.

Prediction performance

Our accuracy did not improve because of the use of pre-trained embeddings, but the model converged faster – allowing us to reduce training from 15 to 5 epochs.

NER Task – Dev set performance





POS Task – Dev set performance