

BiLSTM Tagger - Part 3 Report

1. Training Parameters

For both POS and NER tasks, the BiLSTM tagger was trained for 5 epochs with an evaluation every 500 sentences.

The representation modes tested were:

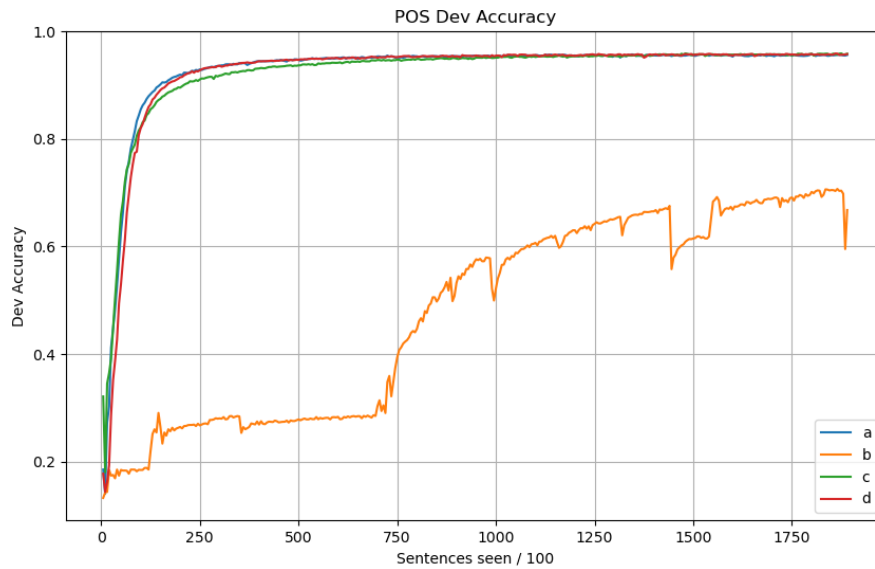
- (a) Word embeddings
- (b) Character-level LSTM
- (c) Word + prefix/suffix embeddings
- (d) Combination of (a) and (b) with a linear layer

Hyperparameters:

- Word embedding dim: 50
- Hidden dim: 100
- Char embedding dim: 30
- Char hidden dim: 50
- Optimizer: Adam
- Learning rate: 1e-3
- Batch size: 64

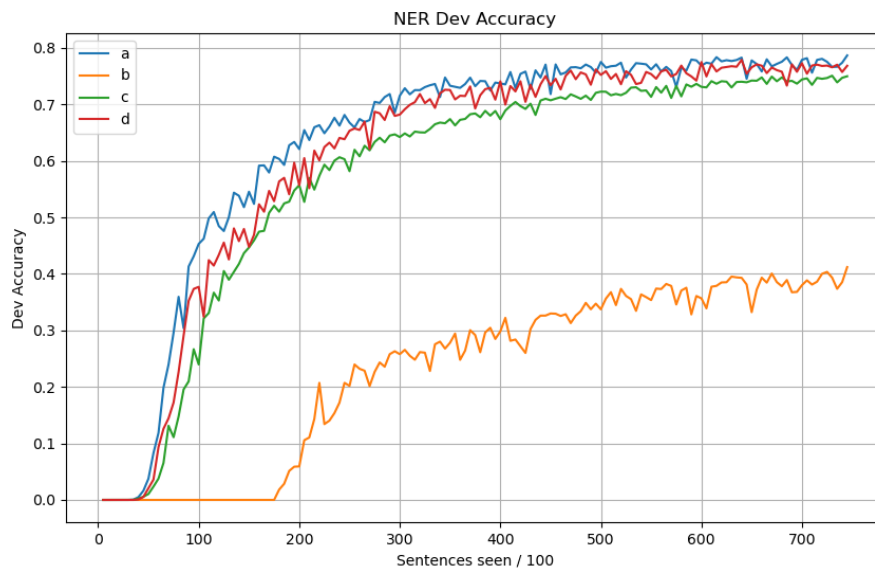
2. POS Dev Accuracy

The plot below shows the development set accuracy across training for each input representation on the POS tagging task.



3. NER Dev Accuracy

The plot below shows the development set accuracy across training for each input representation on the NER task.



4. Observations

- Representation (a), using pretrained word embeddings, performs best overall on both

tasks.

- Mode (b), using only character-level embeddings, performs significantly worse, especially on NER.
- Mode (d), the combination of (a) and (b), closely follows the performance of (a) on both tasks, showing slight improvement in some cases.
- Mode (c), which combines word embeddings with prefix/suffix embeddings, performs worse than (a) and (d) but better than (b).
- For NER, character information adds more value due to its morphological complexity, making mode (d) slightly more competitive.
- For POS, word embeddings dominate, and adding character-level features in mode (d) does not lead to major improvements.