DL4Seq - Assignment 3

Otmazgin, Shon 305394975

Rassin, Royi 311334734

May 15th, 2021

1 BiLSTM

1. Embedding sizes - 50 for both: char and token embeddings.
2. LSTM hidden-size - character-level: 50, token-level: 768.
3. Linear layer - input: 2 * 768, output-size (number of labels).
4. Char-Tokens Linear layer for repr-d - input: 2 * 50, output: 50.
5. Dropout - 0.3, only for the linear layer.
6. Optimizer - AdamW with weight-decay: 0.01 and learning-rate: 1e-3.

1.1 Batching

In order to optimize the training, we implemented mini-batching and padded the sentences according to the longest one in the batch, instead of the entire dataset. This resulted in better overall performance since the network has to work with shorter sequences and less pads, which are semantically meaningless. Further, it of course utilized the GPU in a more efficient manner.

1.2 Results

1.2.1 NER - accuracy

1. a - 74%.
2. b - 83%.
3. c - 83%.
4. d - 86%.

1.2.2 POS

1. a - 95%.
2. b - 96.3%.
3. c - 96.7%.
4. d - 97%.

2 Appendix

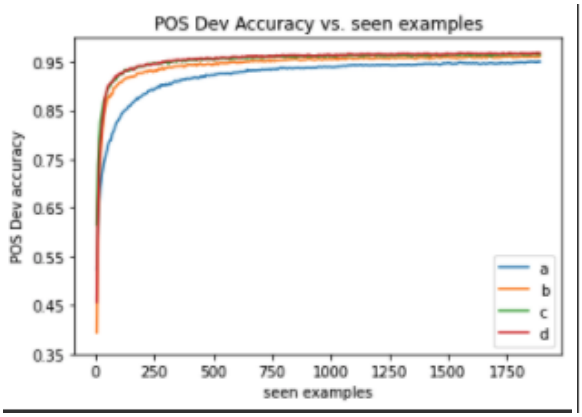

Figure 1: POS - learning curves

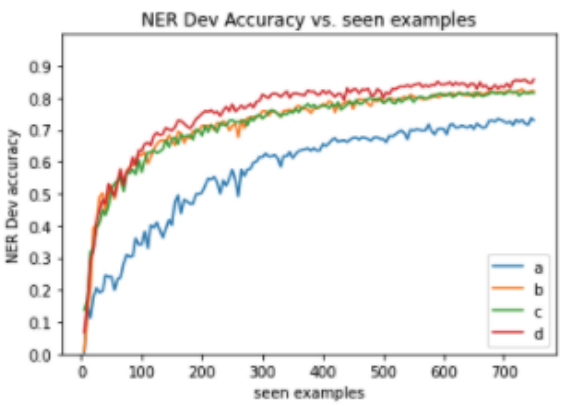

Figure 2: NER - learning curves