Assignment 4

Task 1: Simple Bidirectional LSTM model

What are the precision, recall and F1 score on the dev data?

processed 51577 tokens with 5942 phrases; found: 5715 phrases; correct: 3739.

accuracy: 93.47%; precision: 65.42%; recall: 62.92%; FB1: 64.15

LOC: precision: 76.68%; recall: 80.57%; FB1: 78.58 1930

MISC: precision: 69.82%; recall: 66.49%; FB1: 68.11 878

ORG: precision: 56.89%; recall: 56.67%; FB1: 56.78 1336

PER: precision: 56.40%; recall: 48.10%; FB1: 51.92 1571

Command to run:

python3 task1\_script.py train dev test blstm1.pt

Hyperparameters:

embedding\_dim = 100

lstm\_hidden\_dim = 256

lstm\_layers = 1

lstm\_dropout = 0.33

linear\_output\_dim = 128

batch\_size = 16

learning\_rate = 0.1

num\_epochs = 20

step\_size=10

gamma=0.9

Description:

The code defines a PyTorch dataset class called NERDataset which takes in a list of sentences and their corresponding tags, and converts the words and tags into indices using dictionaries train\_word\_idx and train\_label\_idx. It pads the sequences to have the same length (max\_len) using zeros.

The \_\_getitem\_\_ method returns a tuple of padded sentence and its corresponding tag. The padded sentence is a list of integers, where each integer is the index of a word in the dictionary train\_word\_idx. Similarly, the tag is a list of integers, where each integer is the index of a tag in the dictionary train\_label\_idx.

This dataset can be used to train and evaluate a neural network for named entity recognition (NER) task. The network can take in the padded sentences as input and output the predicted tags for each word in the sentence. The network can be trained using backpropagation and optimization techniques such as stochastic gradient descent (SGD) or Adam. The training process can be monitored using metrics such as f1\_score and accuracy\_score. The network can be evaluated using the test set.

Task 2: Using GloVe word embeddings

What are the precision, recall and F1 score on the dev data?

processed 51578 tokens with 5942 phrases; found: 5751 phrases; correct: 4785.

accuracy: 96.91%; precision: 83.20%; recall: 80.53%; FB1: 81.84

LOC: precision: 93.24%; recall: 81.06%; FB1: 86.72 1597

MISC: precision: 84.89%; recall: 76.79%; FB1: 80.64 834

ORG: precision: 79.98%; recall: 71.51%; FB1: 75.51 1199

PER: precision: 76.80%; recall: 88.44%; FB1: 82.21 2121

Command to run:

python3 task2\_script.py train dev test blstm2.pt

Hyperparameters:

lr=0.5  
step\_size=3  
gamma=0.1  
batch\_size=16  
dimension\_embedding = 100