CSCI 544 – Applied Natural Language Processing

Homework 4 – Vu Truong Si – 6031936649

Python version used: 3.8.8

PyTorch version used: 1.13.0+cpu and gpu

.py file run time: Approximately 4 minutes. Please adjust the number of epochs in model training sections to lower the run time to your liking if code testing is required.

1. Task 1: Simple Bidirectional LSTM model

Firstly, I created a vocabulary from the train file, turning words with less than or equal to 3 occurrences into < unk >. I then included a < pad > token at index 0. This vocabulary is used to encode the sentences in the train, dev, and test files: Each word was mapped to its index in the dictionary. The target labels were also mapped to their indexes. After that, I created Datasets and DataLoaders and prepared for training.

The first BLSTM model was created based on the instruction in the PDF, nothing was changed. The hyperparameters were: Number of epochs = 105, batch size = 32, learning rate = 0.5 and will decrease with momentum = 0.9 when validation loss is no longer changing.

I then saved the trained model and used it for predictions. The results from the dev set are below:

accuracy: 95.59%; precision: 81.03%; recall: 75.97%; FB1: 78.42

LOC: precision: 85.35%; recall: 81.82%; FB1: 83.55 1761

MISC: precision: 81.87%; recall: 72.99%; FB1: 77.18 822

ORG: precision: 74.57%; recall: 67.79%; FB1: 71.02 1219

PER: precision: 80.78%; recall: 77.58%; FB1: 79.15 1769

2. Task 2: Using GloVe word embeddings

For this task, the vocabulary was created based on the Glove file. I also have < unk > and < pad > tokens in the vocabulary at index 1 and 0 respectively. The process of creating Datasets and DataLoaders was exactly the same as the first task.

The model creation part was also the same as the first task. I tuned the hyperparameters slightly but the results did not change much, so I kept everything and increased the number of epochs from 105 to 150. Below are the results for the dev set:

accuracy: 97.69%; precision: 88.10%; recall: 88.93%; FB1: 88.51

LOC: precision: 91.53%; recall: 94.07%; FB1: 92.78 1888

MISC: precision: 83.30%; recall: 80.04%; FB1: 81.64 886

ORG: precision: 79.85%; recall: 80.39%; FB1: 80.12 1350

PER: precision: 92.85%; recall: 94.46%; FB1: 93.65 1874