Question 1:

1. Modify dataset for the implementation “NER using CRF”.

Number distributing of the emerging entity dataset:

Training data: 3394, valid data: 1009, test data: 1287. There are 13 entity labels.

Because the emerging entity dataset do not have POS, we add this feature by nltk toolkit.

I also add several features:

word[-1:], word[1:], word[2:], word[3:]

The final F1 score is 0.135.

B. Modify dataset for the implementation “NER using SoftMax”.

Here I’ve removed the BIO in the labels, so there are 7 entity labels.

Say given a single training data like:

I come from Chicago and like Elon Musk .

The labels like:

O O O Location O O Person Person O

For a batch input, padded labels have been masked.

Three dropout layers are added after the embedding layer, the hidden layer and the output layer to avoid over-fitting.

Use Adam optimizer instead of SGD.

I’ve run for 20 epochs and got 0.014.

C. Optimize the hyper-parameters.

20 epochs.

1. Learning rate

|  |  |
| --- | --- |
| Learning rate | F1 Score |
| 0.5 | 0.017 |
| 0.1 | 0.020 |
| 0.05 | 0.023 |
| 0.01 | 0.037 |
| 0.005 | 0.033 |
| 0.001 | 0.040 |

Learning rate: 0.001

1. Embedding size

|  |  |
| --- | --- |
| Embedding size | F1 Score |
| 8 | 0.014 |
| 16 | 0.018 |
| 32 | 0.029 |
| 64 | 0.019 |
| 128 | 0.033 |

1. Hidden layer size

|  |  |
| --- | --- |
| Hidden layer size | F1 Score |
| 8 | 0.032 |
| 16 | 0.015 |
| 32 | 0.028 |
| 64 | 0.027 |
| 128 | 0.023 |

1. Epochs

|  |  |
| --- | --- |
| Epochs | F1 Score |
| 1 | 0.011 |
| 5 | 0.014 |
| 10 | 0.016 |
| 20 | 0.025 |
| 50 | 0.023 |