Report

Name: Advaith Malladi

Method (English):

For this POS tagging purpose, I considered only continous sequences of length 4. Given past 3 words, I would try to assign the POS tag to the 4^{th} word. I chose this number because most chunks in a sentence are not longer than 4 words. The ambiguity of a POS tag for any word can be resolved by looking at it's chunk. If I were to consider the entire sentence length as input, this would result in too much complexity in the model. Not only that, but as the average sentence length is around 13 and the longest sentence is around 40, it would result in a lot of padding tokens, which results in a lot of junk data. The variance in sentence length is also quite high.

English Corpus:

Scores:

On Training:

- Average accuracy: 0.9806323818897638
- Average F1 Score: 0.9527092632728124
- Average Recall: 0.9535147266101831
- Average Precision: 0.9578111130616309

On Testing:

- Average accuracy: 0.9561941967560694
- Average F1 Score: 0.9157242185409165
- Average Recall: 0.9178236009479254
- Average Precision: 0.9265577653194238

On Validation:

- Average accuracy: 0.9590845360205724
- Average F1 Score: 0.9216825489868614
- Average Recall: 0.9229872282478895
- Average Precision: 0.9311847087549251

Hyperparameters Used:

- epochs: 50
- batch size = 128
- sequence length = 4
- hidden dimension = 256
- layers = 1
- embedding dimension = 256
- hidden2tag layer: input: 4*256, output: target size

Let TP = True Positives
FP = False Positives
FN = False Negatives

Precision = TP/TP+FP
Recall = TP/TP+FN

Analysis:

- From the analysis, it is evident that the model did not underfit as the validation scores are quite close to the training scores.

- The testing and validation metrics are both relatively close in value, indicating that the model is generalizing well to new data.

- The F1 score, which is a harmonic mean of precision and recall, is a good measure of the overall performance of the model. The F1 score is high on all three datasets, which indicates that the model has a good balance between precision and recall.

- The task of POS taggings isn't too complex as we were able to achieve these scores using the above mentioned hyperparameters.

THE END