

Assignment 2

Anshul Tomar 170070007
Parth Shettiwar 170070021
Anwesh Mohanty 170070009

October 18, 2020

Assignment 1 - Finished

All the results, report and the code have been submitted on moodle.

Maximum Entropy Markov Models - Finished

Based on the confusion matrix, the final results obtained are-

Tag	Precision	Recall	F1 Score
B	0.950	0.948	0.949
I	0.923	0.916	0.919
O	0.930	0.958	0.944

The overall accuracy comes out to be 93.78%.

	Precision	Recall	F1 Score
Overall	0.9376	0.9379	0.9378

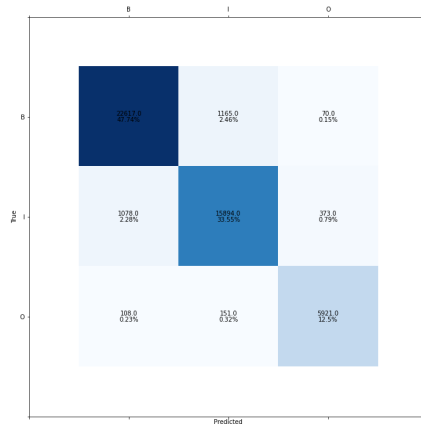


Figure: Confusion Matrix for MEMM

Bidirectional Long Short-Term Memory - Finished

Based on the confusion matrix, the final results obtained are-

Tag	Precision	Recall	F1 Score
B	0.9692	0.9663	0.9677
I	0.9472	0.9521	0.9497
O	0.958	0.955	0.9565

The overall accuracy comes out to be 95.96%.

	Precision	Recall	F1 Score
Overall	0.9598	0.9597	0.95971

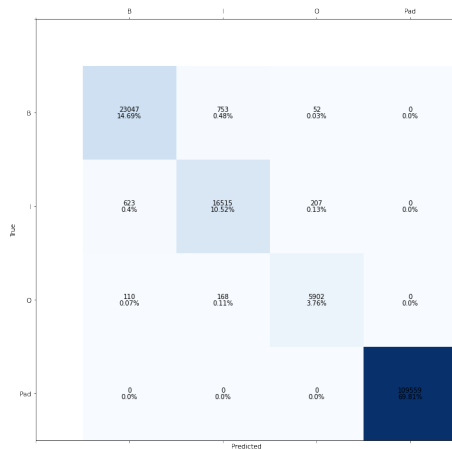


Figure: Confusion Matrix for BiLSTM

Conditional Random Fields - Finished

From this matrix, we can easily calculate the various accuracies required.

Tag	Precision	Recall	F1 Score
B	0.944	0.945	0.9445
I	0.917	0.915	0.916
O	0.951	0.949	0.950

	Precision	Recall	Accuracy
Overall	0.9352	0.9347	0.9348

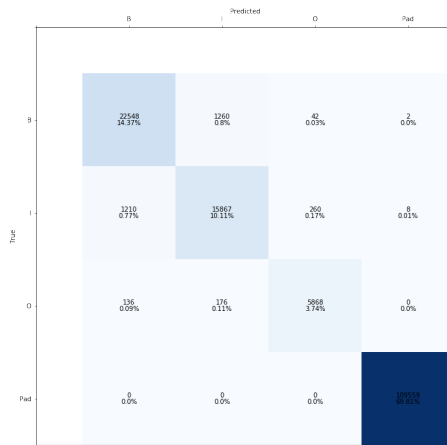


Figure: Confusion Matrix for CRF

Error Analysis

- All the three tags i.e 'B', 'I' and 'O' are predicted quite accurately by the model after just training for 10 epochs with almost greater than 95% accuracy for each tag. This makes it the best model out of the three models taken for comparison.
- The overall accuracy is closer to the 'B' tag due to more occurrences of 'B' tag. The accuracy of 'I' is slightly lower than that of 'B' which might be due to its reduced instances in the dataset compared to 'B'.
- The highest number of false positives and false negatives are due to mis-classification of 'B' as 'I' and vice-versa. And since the number of 'O' tags present are much lower compared to 'B' and 'I', the mis-classification of 'O' does not affect accuracy that much.
- Since we introduce padding into our model we also have to consider the change in accuracy introduced due to padding. But as can be seen from the confusion matrix, ALL the padding tags have been classified with 100% accuracy with no false positives or negatives. This indicates that the model is completely robust to the padding tags and is not affected by such tags.