REPORT

Roll Number: 2018101052

The following statistics were observed for the following 6 files:

1. **2018101052_LM_train.txt: 49.98745243951391**
2. **2018101052_LM_test.txt: 203.65032722708509**
3. **2018101052_MT1_train.txt: 0.24166411983944842**
4. **2018101052_MT1_test.txt: 0.1568583977781715**
5. **2018101052_MT2_train.txt: 0.037470286710053706**
6. **2018101052_MT1_test.txt: 0.0312297402513026**

For language model files, the average perplexity was better for the train set when compared with the test set.

For the language model developed from scratch, the train set corpus-BLEU score was around 24 whereas for the test set corpus-BLEU score was only around 15.

We observed that the BLEU-score for the model developed using the embeddings from the first part has a very low BLEU-score of around 3 for both train and test.

The model used for training the language model is a LSTM model. The input to the model is a four-gram and the output is the predicted next word.

The model used for training the machine translation model is a time distributed LSTM model with 1024 nodes. The input to the model is a

sentence and the output is the corresponding French sentence. We apply padding to the sentences so that the length of all sentences is uniform.