Model Selection

This file contains some more information regarding how the model is selected and experiments performed for hyper-selection. For many experiments, just the validation accuracy is available as I was **not saving** train and validation loss curves for **initial experiments**. Training for experiments mentioned below are done only for 15 epochs and accuracy on validation set is calculated, if the accuracy after 15 epochs were greater or comparable to accuracy of final model after 15 epochs on validation set then model was trained for longer. These experiments were only performed once and therefore do not account for variance in accuracy. Also the hyperparameters are considered orthogonal, which is not true (e.g. when increasing the number of layers, we should also increase dropout probability) .So the result could be improved by tuning all the parameters together. Learning rate, batch size for learning these models were fixed and not tuned for particular models. Details for selecting the learning rate, batch size are mentioned in the main report.

Embedding Size: 3 values of hyper-parameter were tested.

Size	Accuracy
200	76.1
250	75.83
300	76.32

We can see that embedding size of 250 has worse performance which is not justified, this is because these experiments were only performed once and there is variance in accuracy therefore embedding size 200 has better performance than embedding size 250.

Hidden State dimension: 4 values of hyper-parameter were tested.

Size	Accuracy
200	75.56
256	76.32
300	75.19
350	76.61

Hidden State of 350 performs best, but the final difference between model with hidden state dimension of 256 and model with hidden state dimension of 350 was very less, so model with hidden state dimension 256 was chosen.

Hidden layers in Fully connected network: 3 values of hyper-parameter was tested.

Size	Accuracy
2	75.07
3	76.32
4	75.14

We see that accuracy decreases using 4 hidden layers, but this may be due to overfitting and could probably be avoided by increasing dropout probability.

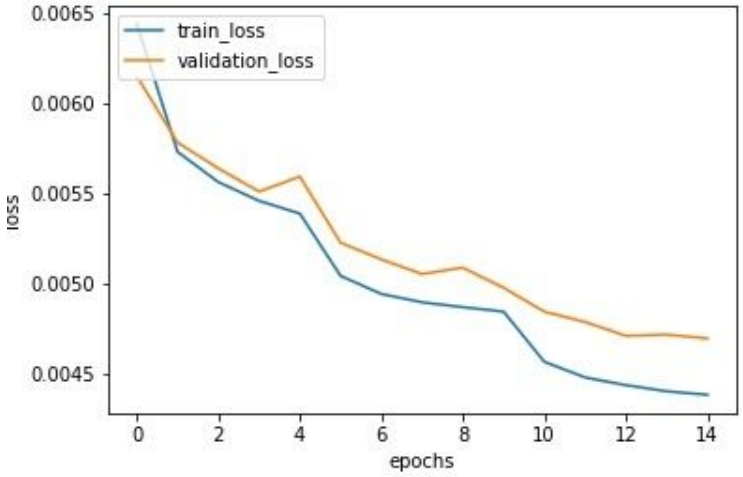
Learning rate, Batch size : Details are mentioned in the main report.

LSTM or Bi-LSTM:

Method	Accuracy
LSTM	74.15
Bi-LSTM	76.32

We can see that Bi-LSTM works better than LSTM which is expected.

Train and Validation Loss graphs for some experiments:



Fig, Embedding dimension 200

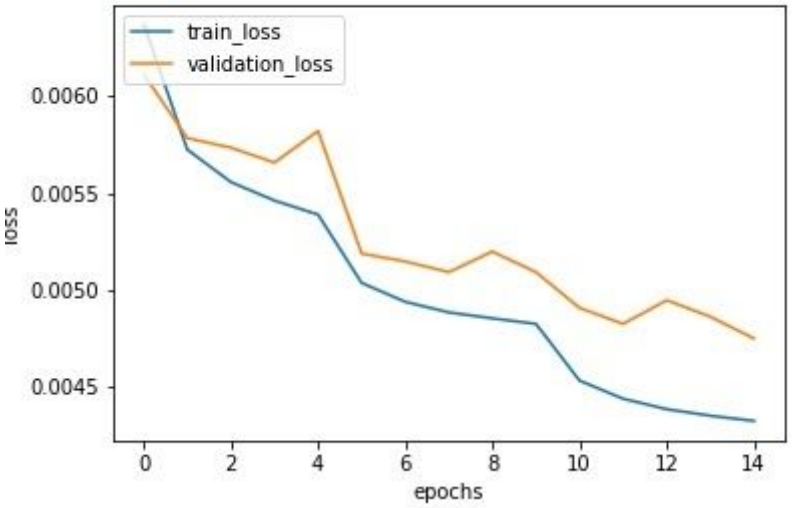


Fig. Embedding dimension 250

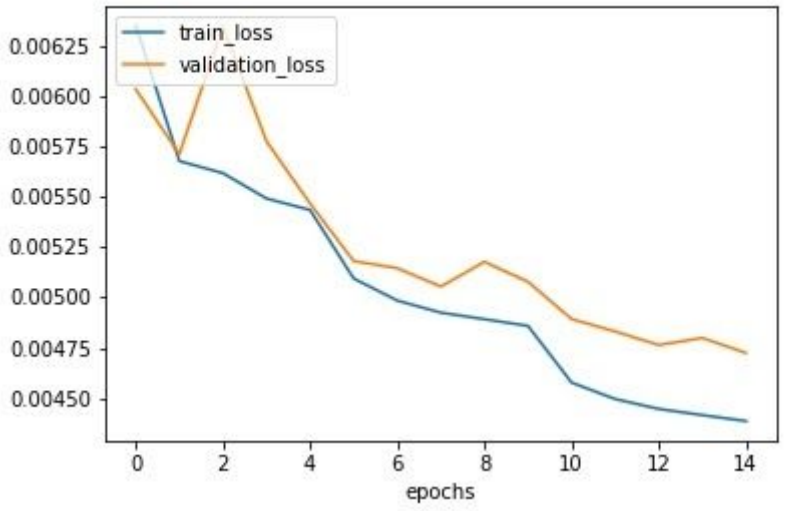


Fig. Hidden state dimension 200

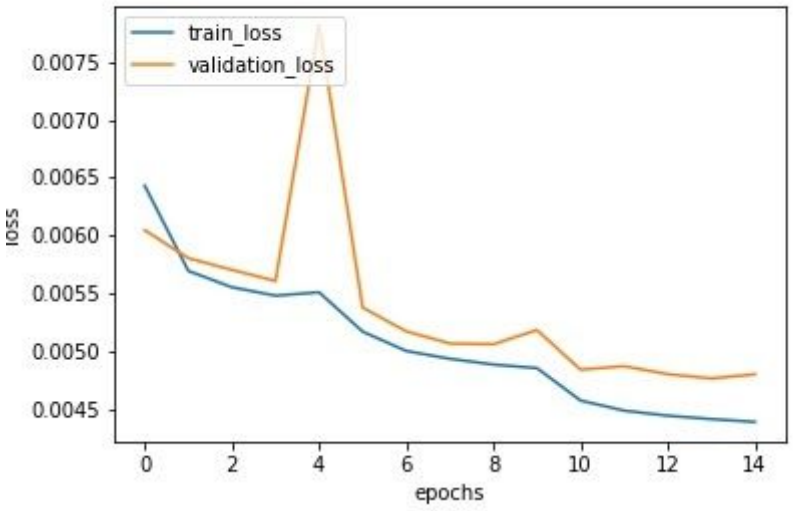


Fig. Hidden state dimension 300

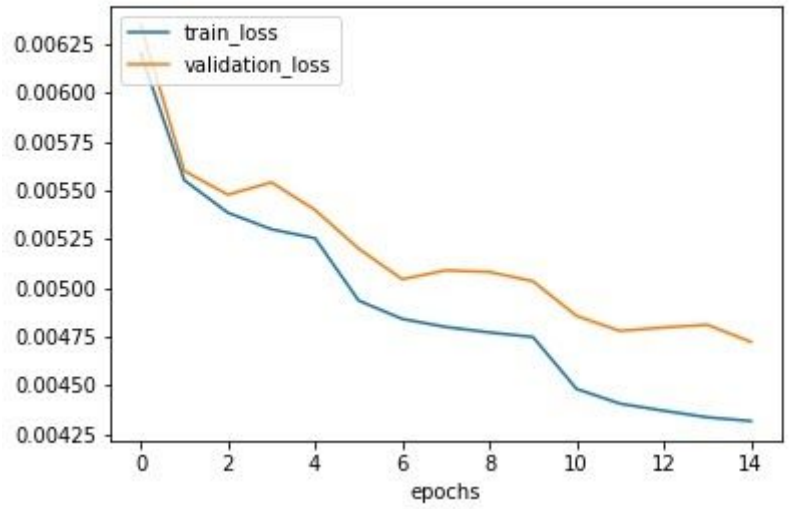


Fig. Hidden layer=2 in FC layers

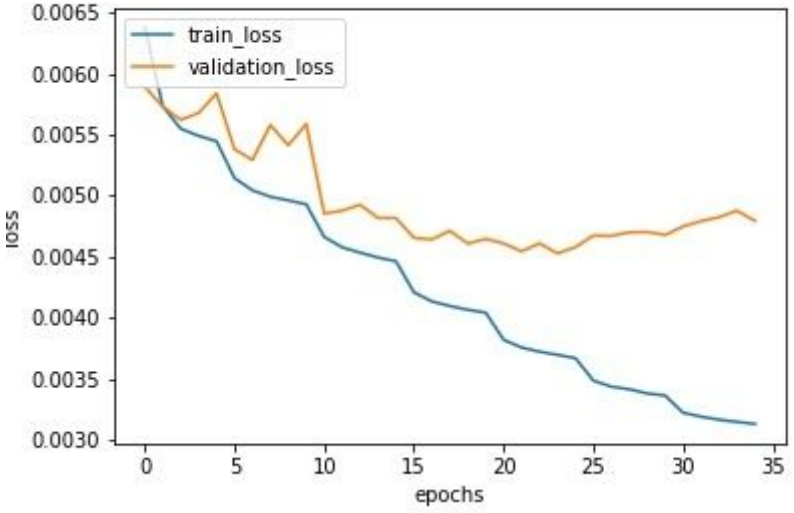


Fig. Hidden layer=3, Embedding dim=300,
Hidden state dim=256