Sentence Classification using Densely Connected Bidirectional LSTM

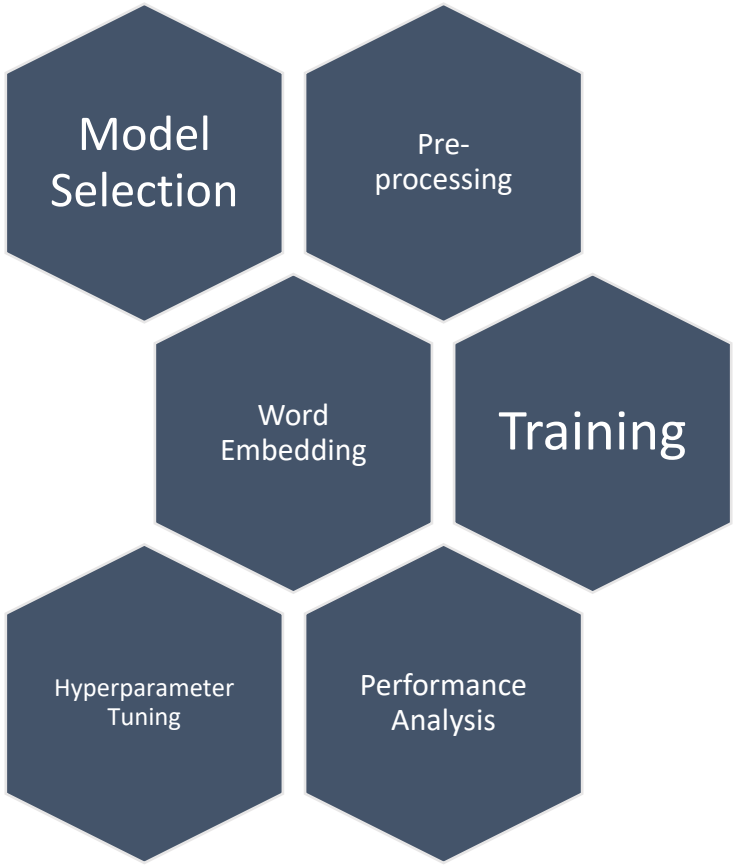
Md Tahmid Yasar

Khalid Saifullah

Overview

Model

1) Bi-LSTM
2) Deep Stacked Bi-LSTM
3) Densely Connected Bi-LSTM



Model Details

Hyperparameters	Value
Learning rate	0.01
Learning rate decay	0.05
Optimizer	Adam
Clip value	5
Dropout keep probability	0.5

Dimensions	Bi-LSTM	DS-Bi-LSTM	DC-Bi-LSTM
Number of layers	1	5	15
Number of units, up to penultimate layer	0	39	13
Number of units in last layer	300	100	100
Number of parameters	1582306	392834	1445554

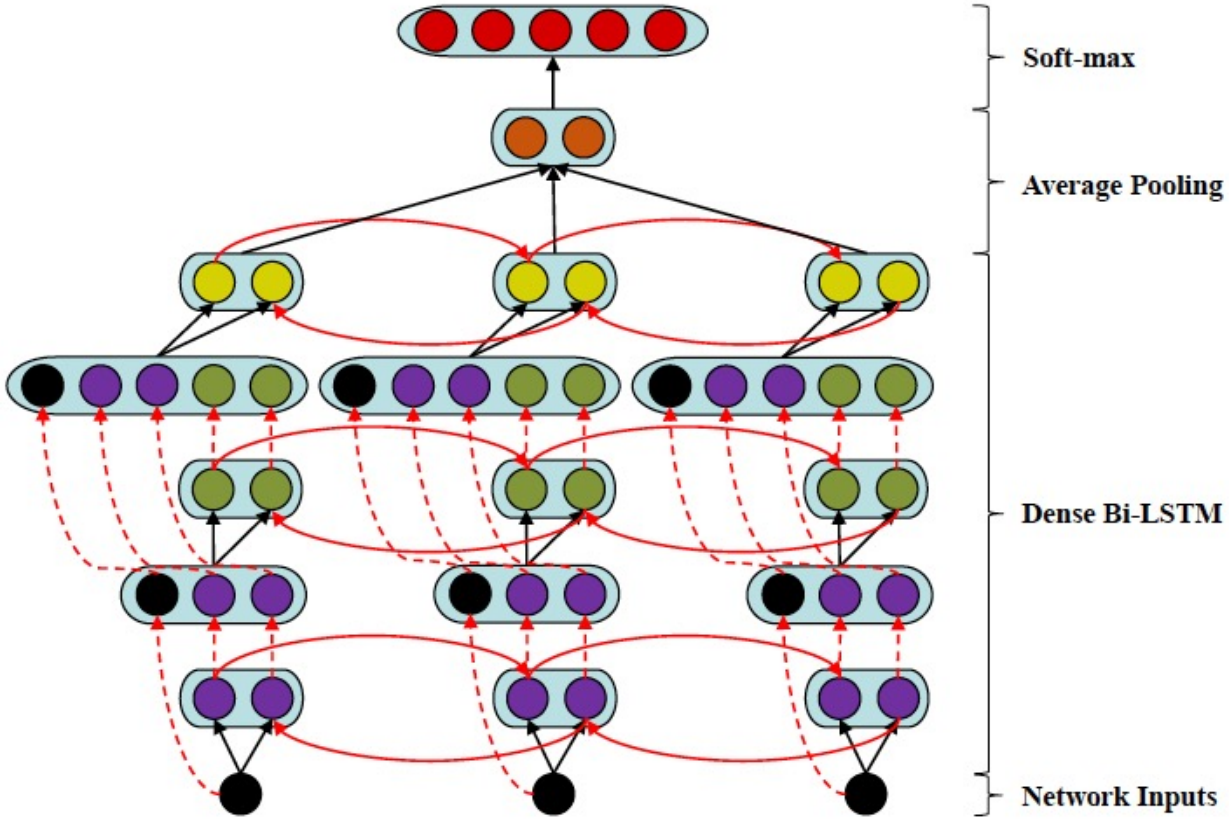


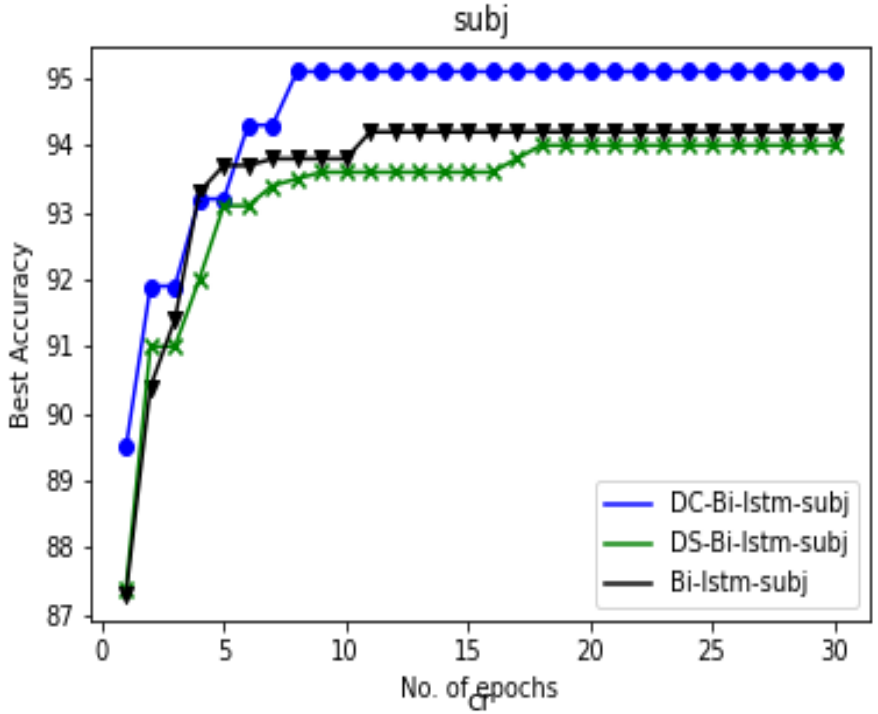
Illustration of Densely Connected Bi-LSTM

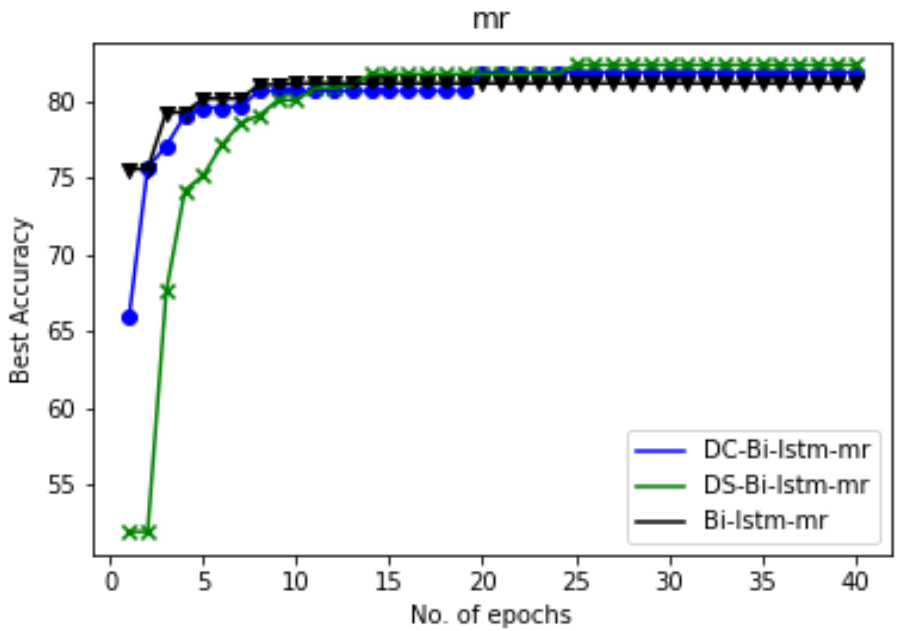
Performance Evaluation

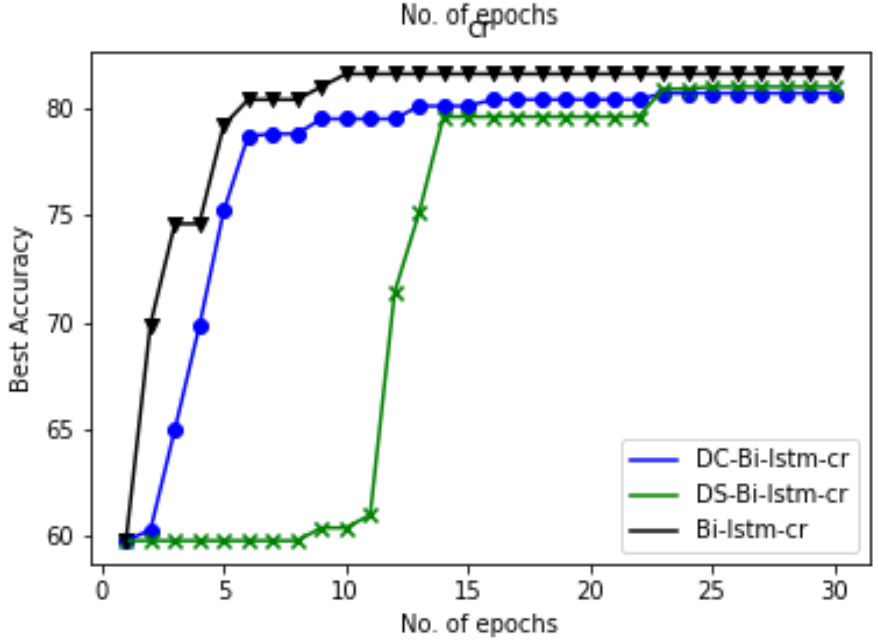
Comparative performance analysis of maximum accuracy and corresponding epochs

Model	MR	SST-1	SST-2	Subj	TREC	CR	MPQA
Bi-LSTM	81.2 (8th epoch)	46.4 (8th epoch)	86.5 (14th epoch)	94.2 (8th epoch)	94.8 (20th epoch)	81.6 (10th epoch)	87.3 (17th epoch)
DS- Bi-LSTM	82.4 (25th epoch)	46.0 (33rd epoch)	86.5 (30th epoch)	94.0 (18th epoch)	91.3 (30th epoch)	81.0 (25th epoch)	87.2 (21st epoch)
DC-Bi-LSTM (ours)	81.9 (20th epoch)	47.2 (23rd epoch)	87.5 (12th epoch)	95.1 (8th epoch)	94.7 (15th epoch)	80.7 (23rd epoch)	86.7 (6th epoch)
DC-Bi-LSTM (paper)	82.8	51.9	89.7	94.5	95.6		

Experimental Results







Discussion and Conclusion

- Implementation of 3 LSTM based models
- Gradient vanishing and overfitting issues
- Performance evaluation on 7 datasets
- Poor performance of DS-Bi-LSTM with high number of parameters
- Validation of the claim of reference paper with our experiment