Learning sentence representations from natural language inference data

Konstantin Todorov, No. 12402559
Statistical Methods for Natural Language Semantics
University of Amsterdam

Github repository link - https://github.com/ktodorov/uva-semantics-19

For all models we can observe that hypotheses where almost all of the sentence is similar to the premise the model predicts correctly as entailment but if the differences in the hypothesis become more complex then the models fail to predict correctly

1 Mean encoder

- test macro accuracy: 61.4346%
- test micro accuracy: 61.4413%
- SentEval

	devacc	**acc**	**ndev**	**ntest**
MR	77.87	76.63	10685	10685
CR	80.17	78.47	3775	3775
SUBJ	91.16	91.14	10021	10021
MPQA	87.51	87.57	10606	10606
TREC	73.59	82.6	5452	500
SST2	79.59	79.68	872	1821

2 Uni-LSTM encoder

- test macro accuracy: 33.7865%
- test micro accuracy: 33.7438%

3 Bi-LSTM encoder

- test macro accuracy: 34.4460%
- test micro accuracy: 34.4258%

4 Bi-LSTM with max-pooling encoder

- test macro accuracy: 37.6724%
- test micro accuracy: 37.6832%