

Assignment 1: InferSent

Shantanu Chandra
University of Amsterdam
Student ID: 12048461
shantanu.dechandra@student.uva.nl

1 Training process

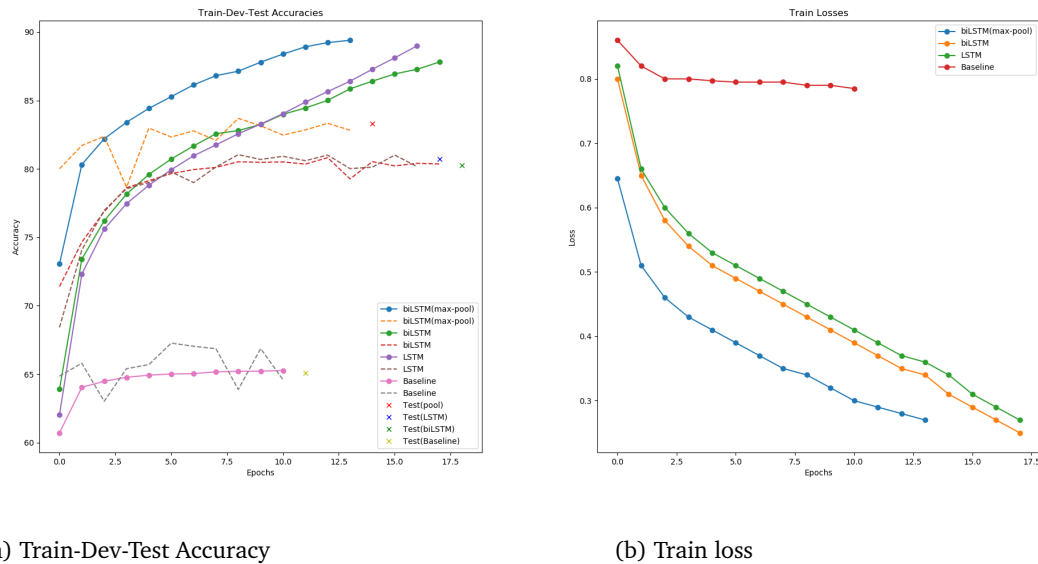


Figure 1: Test accuracy achieved by all the models: **Baseline** = 65.07%, **LSTM** = 80.73%, **biLSTM** = 80.26%, **biLSTM(max-pool)** = 83.32%

2 Results and Analysis

Model	MR	CR	SUBJ	MPQA	SST	TREC	MRPC	SICK-R	SICK-E	STS-14
Baseline	77.19	78.14	91.1	87.85	80.29	83.0	72.87/81.38	0.80	78.49	0.54/0.55
LSTM	70.15	74.38	82.25	85.39	74.03	58.8	72.0/81.66	0.83	81.96	0.49/0.48
biLSTM	74.4	78.6	89.28	87.88	78.14	83.6	73.16/81.98	0.87	84.6	0.55/0.57
biLSTM(max-pool)	78.08	81.35	92.24	88.85	82	88.4	74.38/81.44	0.88	85.26	0.64/0.65

Table 1: SentEval scores on 10 downstream tasks of the 4 models. biLSTM with max-ppol outperforms all the other models at all the 'transfer' tasks.

Model	NLI dev	test	Transfer macro	micro
Baseline	64.6	65.07	81.1	82.73
LSTM	80.73	80.0	75.87	75.65
biLSTM	80.26	79.83	81.21	82.65
biLSTM(max-pool)	83.82	83.83	84.82	85.18

Table 2: Macro- and Micro- Accuracies on SentEval scores ('transfer task') of the 4 models along with training performance measures ('NLI task').

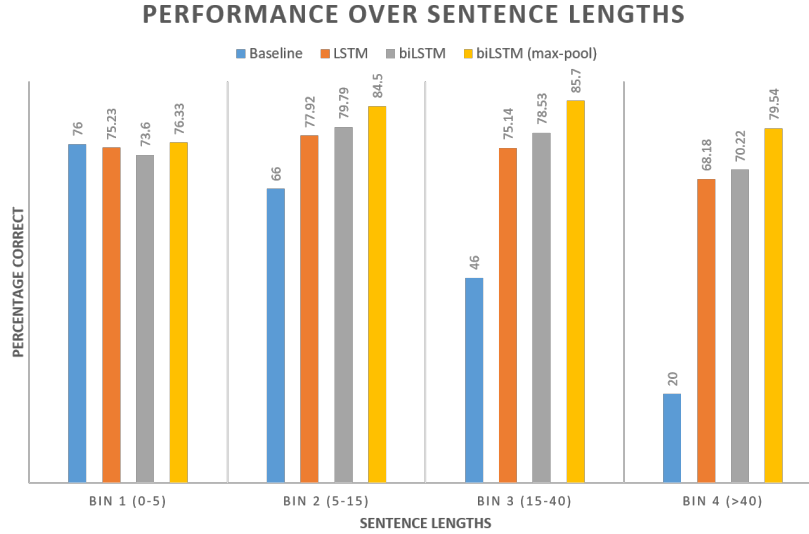


Figure 2: Performance of the models over different ranges of the sentence lengths. We can observe that both biLSTM models perform much better than others as the sentence length increases. For shorter sentences, performance of all the models is comparable.

A Appendix

Premise: Excellent

Hypothesis: Pathetic

Base = [9.6518, 2.8749, -12.3981] (*Entailment*)

LSTM = [-1.7245, 0.5521, 1.1724] (*Contradiction*)

biLSTM = [-4.9323, 1.2784, 3.6504] (*Contradiction*)

biLSTM(max-pool) = [-15.0683, 1.5829, 13.4708] (*Contradiction*)

Premise: This burger is very good

Hypothesis: This burger is very bad

Base = [2.3129, 2.1854, -4.4383] (*Entailment*)

LSTM = [-0.0933, 0.3281, -0.2348] (*Neutral*)

biLSTM = [-2.6843, -0.1019, 2.7698] (*Contradiction*)

biLSTM(max-pool) = [-3.0693, 0.2361, 2.8359] (*Contradiction*)

Premise: I am a boy

Hypothesis: I am **not** a boy

Base = [2.8127, 0.7847, -3.5601] (*Entailment*)

LSTM = [0.4299, -0.1724, -0.2575] (*Entailment*)

biLSTM = [0.4546, -0.3356, -0.1364] (*Entailment*)

biLSTM(max-pool) = [-0.8157, -1.1039, 1.9276] (*Contradiction*)

Premise: The **man** is riding a bike wearing a blue helmet

Hypothesis: The **woman** is riding a bike wearing a blue helmet

Base = [2.0862, -0.4986, -1.5574] (*Entailment*)

LSTM = [-0.5291, -0.5784, 1.1075] (*Contradiction*)

biLSTM = [-1.6187, -1.0998, 2.6845] (*Contradiction*)

biLSTM(max-pool) = [-7.2812, -1.5112, 8.7917]] (*Contradiction*)

Premise: A lady is in the park

Hypothesis: A lady is in the house

Base = [0.9784, 0.1472, -1.0930] (*Entailment*)

LSTM = [-1.2440, -0.5133, 1.7574] (*Contradiction*)

biLSTM = [-3.6487, -1.1938, 4.8015] (*Contradiction*)

biLSTM(max-pool) = [-4.7834, -2.0677, 6.8549] (*Contradiction*)

Premise: I tried to make this a really long sentence but I am failing so bad that I have to write something that does not even make sense

Hypothesis: This is a short sentence

Base = [1.9736, 1.4264, -3.3548] (*Entailment*)

LSTM = [0.3392, 0.2548, -0.5939] (*Entailment*)

biLSTM = [-4.6421, 1.4537, 3.1921] (*Contradiction*)

biLSTM(max-pool) = [1.4613, 0.5478, -2.0038] (*Contradiction*)

Premise: I can tell you that I do not like fast cars and also I can say for sure that I do not like bikes

Hypothesis: I can tell you that I like fast cars and also I can say for sure that I do not like bikes

Base = [2.2253, 0.8985, -3.0780] (*Entailment*)

LSTM = [1.6140, 0.0471, -1.6611] (*Entailment*)

biLSTM = [0.9712, 0.2159, -1.1899] (*Entailment*)

biLSTM(max-pool) = [0.5556, -0.0174, -0.5308] (*Entailment*)

Premise: I can tell you that I do not like fast cars and also I can say for sure that I do not like bikes

Hypothesis: I can tell you that I like fast cars and also I can say for sure that I like bikes

Base = [2.2253, 0.8985, -3.0780] (*Entailment*)

LSTM = [1.1728, 0.0794, -1.2522] (*Entailment*)

biLSTM = [-0.1934, 0.0876, 0.0975] (*Contradiction*)

biLSTM(max-pool) = [-0.7368, -0.1455, 0.8880] (*Contradiction*)