
Practical 1: Learning Sentence Representations

Steven van de Graaf

12234036

steven.vandegraaf@student.uva.nl

All code and output for this assignment are available [here](#).

While all coding is done (perhaps with the exception of some minor bug fixes as they come up during training), the training is not. I have managed to train the Baseline model for the full 20 epochs, but, as of yet, have been unable to train the LSTM, BiLSTM or the MaxBiLSMT models. While I did try to make use of SURFsara's LISA computing cluster to train my models, ever since the afternoon of Monday 22 April, LISA has not run anything for me.

Baseline

Below, you'll find a plot of the accuracy and loss curves obtained during training of the Baseline model in Fig. 1.

As you can see, the accuracy obtained during training on the DEV set (as well as the TRAIN set) is 0.333. With $n = 3$ classes, this amounts to random guessing. Also note that the losses are both constant and extremely low. This is indicative of something fishy!

Nonetheless, I evaluate this model on 3 transfer tasks from the SentEval framework, namely: MR, CR and MPQA, the results of which are tabulated below in Table 1.

As you can see, these transfer task results are quite good. In fact, they're highly comparable to the GloVe BOW results listed in Table 4 of Conneau et al. [2017]. This makes sense, as the Baseline model is just that; a BOW model using GloVe features.

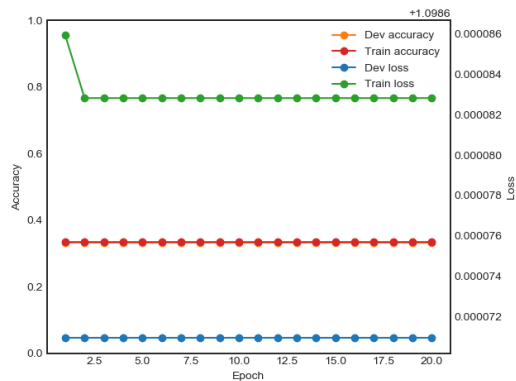


Figure 1: Accuracy and loss curves obtained for the Baseline model

| Task | Accuracy |
|------|----------|
| MR | 77.67% |
| CR | 79.84% |
| MPQA | 87.43% |

Table 1: Results obtained on some SentEval tasks for the Baseline model

References

Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*, 2017.