

NLU course projects

Amir Gheser (247173)

University of Trento

amir.gheser@studenti.unitn.it

1. Introduction (approx. 100 words)

With the purpose of obtaining a suitable model for detection of aspects for aspect based sentiment analysis, I fine tuned BERT uncased on the Laptop partition of the SemEval2014 task4 [?]. This dataset is a collection of reviews labelled with the following labels:

- 'O': Not an aspect
- 'T-POS': Positive aspect
- 'T-NEG': Negative aspect
- 'T-NEU': Neutral aspect

The model I trained was aimed only at identifying aspects; hence, during parsing of the dataset samples were relabelled as not an aspect (1) or aspect (2).

2. Implementation details (max approx. 200-300 words)

SemEval provided a training and test set. As the datasets were quite small I did not split the training set but used the test set as validation set as well. As BERT encodes using a byte-pair algorithm, samples had to be adjusted to take padding into account. Training was done over a maximum range of 20 epochs, although most training runs were aimed at stopping in 10 epochs. The training loss of choice was Cross Entropy, set to ignore 'O' labels.

I trained the model with AdamW with an initial learning rate of $2e-5$ and reducing it on Plateau by a factor of 0.25 and patience 1.

Padding was removed from labels, and 'O' predictions were always treated as incorrect.

3. Results

I trained the model for 5 epochs as it often revealed to be sufficient. To reduce instability I increased the dropout probability after the linear layer to 0.5 which made curve smoother. And started achieving slightly better F1-scores but still overfitting. The best model was trained with lower learning rates ($1e-5$) and a dropout value of 0.2 as higher dropout values required higher learning rates but weren't yielding satisfactory results.

lr	dropout	F1
$1e-5$	0.1	0.64
$3e-5$	0.1	0.72
$1e-5$	0.2	0.68

Table 1: F1-scores of plotted models

4. References

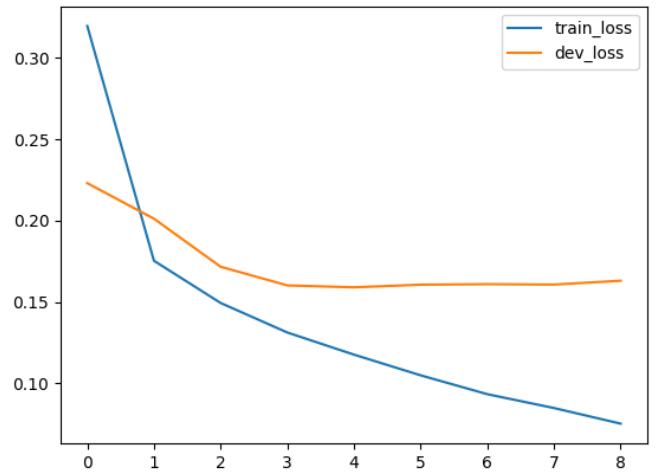


Figure 1: The plot of a model trained over 8 epochs with AdamW and a learning rate of $1e-5$ and dropout 0.1

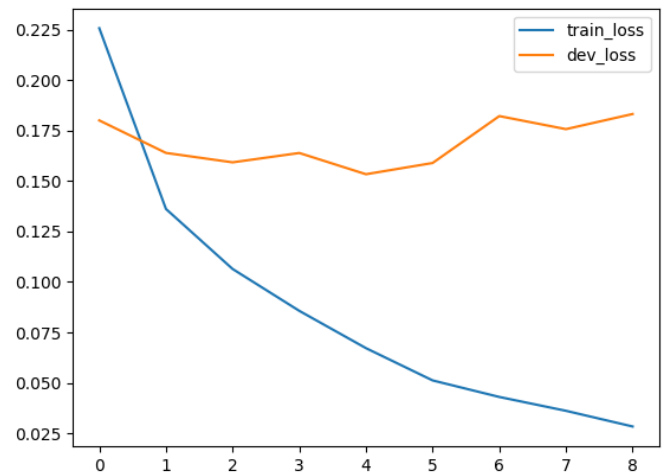


Figure 2: The plot of a model trained over 8 epochs with AdamW and a learning rate of $3e-5$ and dropout 0.1

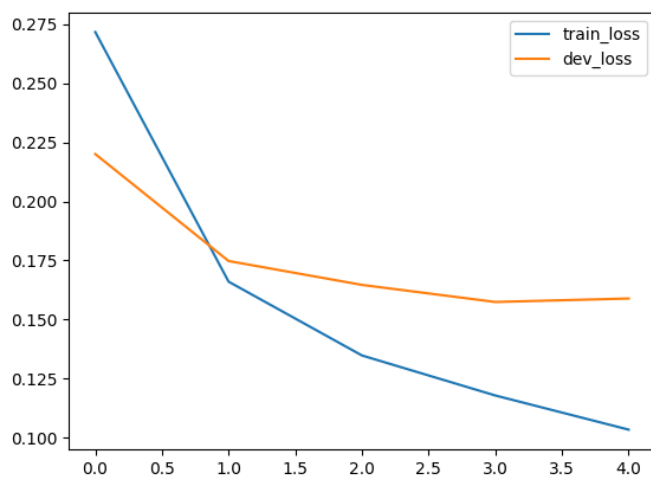


Figure 3: *Model trained with AdamW, dropout 0.2 and learning rate of $1e-5$.*