

NLU course projects: lab 6

Simone Roman (mat. 247181)

University of Trento

simone.roman@studenti.unitn.it

1. Introduction

This assignment focuses on implementing the task of Aspect-Based Sentiment Analysis (ABSA) using a model based on the pre-trained BERT language model. The aim is to accurately extract aspect terms from the given text. Evaluation metrics include F1 score, precision and recall. My approach involves adapting the existing model and code developed for intent classification and slot filling (from the second assignment) to suit the requirements of extracting aspect terms.

2. Implementation details

The idea for the implementation is to adapt the code developed for the second assignment, as the aspect extraction task can be considered a specific case of the slot filling task.

The first change involved handling the dataset. Since the SemEval2014 dataset is different, I transformed each item into a dictionary containing a sentence and its corresponding aspect terms. Each word in the sentence and aspect term was processed, assigning "O" to indicate the absence of an aspect and "T" to indicate the presence of an aspect term. Afterward, I created a map that associates the elements with IDs using the BERT tokenizer. I adjusted the lengths to ensure that each aspect term was correctly aligned with the tokens generated by BERT's tokenizer.

The model architecture required minimal changes. I used BERT with an added single linear layer to perform aspect extraction.

For the evaluation phase, before calculating F1, Recall and Precision, I prepared the references and the correlated hypotheses generated by the model. I removed the "pad" tokens used for alignment with BERT's tokenizer and those added at the end to equalize the batch element lengths.

To calculate the metrics (F1, Precision, and Recall), I modified the provided script [1]. Specifically, I updated the `evaluate_ote` function, which takes the gold standard labels (reference) and the predicted labels (hypothesis) from the model as input. The function counts the true positives (T labels correctly predicted as T), the total T labels in the reference, and the total T labels in the prediction. Using these counts, it calculates Precision, Recall, and F1 score, considering only the aspect terms ('T').

3. Results

The results obtained are shown in Table 1, the metrics include Recall, Precision and F1 Score. Additionally, an image of the training and validation loss is presented in Figure 1.

4. References

- [1] Lixin, "E2e-tbsa: End-to-end target-based sentiment analysis," <https://github.com/lixin4ever/E2E-TBSA/blob/master/evals.py>.

Model name	Recall (%)	Precision (%)	F1 Score (%)
model_11	85.17	88.06	86.59

Table 1: Performance metrics for aspect term extraction using the BERT-based model. The metrics include Recall, Precision and F1 Score.

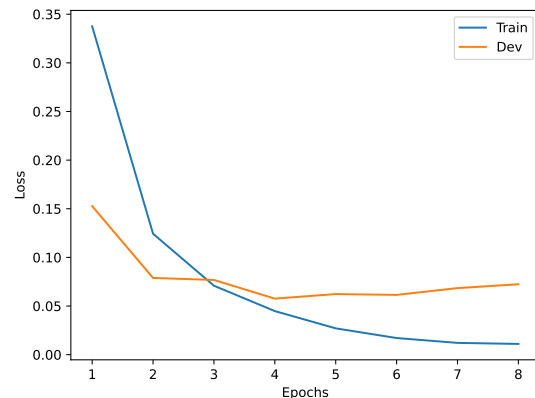


Figure 1: Training and validation loss over epochs. This plot shows the progression of the loss function during the training and validation phases, indicating the model's learning behavior.