

NLU course project lab6

Mattia Rigon (mat. 217182)

University of Trento

mattia.rigondz@studenti.unitn.it

1. Introduction

In this project a model based on Bert was implemented for the Aspect Based Sentiment Analysis task regarding the extraction of the aspect term only.

The Laptop partition of SemEval2014 task 4 was used as dataset, and as evaluation metrics the F1 score, Precision and Recall were used.

2. Implementation details

First of all, a data loader for the "Laptop partition of SemEval2014 task 4" dataset was implemented. The dataset appears in a text format and offers two versions: raw and pre-processed. The pre-processed version was preferred in order to have a better training. In the dataset there are sentences and tagging scheme. For this project there are 3 tagging schema: the T which stays for target, the O which stays for others and padding class. Since the classes are just 3, a particular mapping slot to id was not implemented, but it was decided that 0 stays for padding, 1 for "Target" and 2 for "Other".

Let's discuss the architecture of the model, which is composed of a BERT pre-trained layer (Devlin et al. [1]) and a linear layer in order to fine-tune BERT on the task of sequence tagging. In order to regularize the model a dropout layer was applied before the linear one.

As in the assignment of the lab5, a pre-process to the data was necessary in order to placed as input in the BERT layer. The problem was the mismatch between the length of the tokenized sentences introduced by the BERT tokenizer and the aspects provided for that sentences. As for the other assignment of the lab5, in the slots ground truth the padding token was introduced for all the sub-tokens that star with "##". In this way the slot was assigned only to the first subtoken that the BERT tokenizer had generated for that word.

The criterion loss used is the cross entropy loss.

At the evaluation stage, the padding tokens that had been introduced in the pre-process phase were deleted and all the hypothesis aspects placed in the same indexes of the deleted pad were eliminated as well. Table 1 clarifies this step.

Table 1: *Deleting process in the evaluation of the pad token from the ground truth, and the slot predicted in the same position in the hypothesis.*

reference	O	T-POS	{PAD}	O	O
hypothesis	O	T-POS	Ø	T-NEG	O

Once that both the reference and hypothesis had been post-processed, they passed through the evaluation process which was slightly modified in order to be adapted to our data.

3. Results

As stated by Hu et al. [2] the use of sequence tagging in this task provides good results. The model was trained for 10 epochs by using Adam as optimizer with a learning rate of $5 * 10^{-5}$, the batch size was set to 64 for training and 128 for test.

Table 2: *Evaluation Metrics*

Metric	Score
Precision	0.86
Recall	0.88
F1 score	0.87

4. References

- [1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.
- [2] M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv, "Open-domain targeted sentiment analysis via span-based extraction and classification," 2019.