

# NLU course projects - Lab 6 (SA)

Matteo Mascherin (247183)

University of Trento

matteo.mascherin@studenti.unitn.it

## 1. Introduction

Aspect based sentiment analysis is a NLU task which consists in the extraction of special words called aspects. After aspects are obtained sentiment analysis can be performed on the parsed sentence. As shown by [1], a Bert model can be used and fine tuned to perform the aspect extraction from a corpus. The corpus selected is the Laptop partition of SemEval2014 provided by a project implementing the same paper [2]. The project is evaluated considering precision, recall and f1 metrics on the extracted aspects. A version of the evaluation script is available at this link [3].

## 2. Implementation details

The neural architecture relies on a pretrained Bert model which output is used to feed a linear layer. The last hidden state for each token is used to determine the aspect identification that is performed using a linear layer right after Bert is called.

### 2.1. Tokenization

In order to load the dataset all the sequences and the aspect tags are collected inside a dictionary. Then every word is tokenized and in case is splitted the additional subtokens are padded inside the aspects array (2 for clarification). For this task three labels are used to differentiate between the tags:

- **O - Other:** to indicate a normal token
- **T - target:** identify an aspect token
- **PAD - padding:** for subtokens

### 2.2. Evaluation

For the evaluation part, precision, recall and f1 are computed using the script proposed by [3]. Before computing the metrics, all token corresponding to "PAD" in the aspects tags are removed from every sequence. In the script in order to count the number of hit or correct tagging, tokens corresponding to "O" tag in the predicted aspects are not considered. The same thing is done for the ground truth aspects to compute the above cited measures.

### 2.3. Training settings

As optimizer during training AdamW was used with a learning rate of  $1e-4$ . Since the model is already pretrained,  $\approx 30$  epochs are enough to reach convergence. For better generalization capabilities a dropout with probability 0.1 is inserted inside Bert.

## 3. Results

The aspect tagging performance is measured in term of precision, recall and F1 score. The result are shown in 1. Analyzing the outcome of the training process, it can be observed that Bert is able to reach great numbers in term of F1 score. Furthermore,

Model	Laptop14		
	Precision	Recall	F1
Frozen Bert	77.2	31.4	44.6
Bert	<b>83.5</b>	<b>66.0</b>	<b>73.7</b>

Table 1: Results in term of precision, recall and f1 score for aspect tagging

utterance	my	net	##book	does	not	work
aspects	O	O	PAD	O	O	O

Table 2: Example of sequence tokenization

some parameters changing were tried, but since the dataset is pretty small the best result achievable are the one shown. In the task the evaluation method has a huge impact on the numbers. This is caused by the fact that all the tokens predicted as "O" are discarded from the number of matching.

## 4. References

- [1] M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv, "Open-domain targeted sentiment analysis via span-based extraction and classification," 2019.
- [2] "Dataset." [Online]. Available: [https://github.com/lixin4ever/E2E-TBSA/tree/master/data\\_conll](https://github.com/lixin4ever/E2E-TBSA/tree/master/data_conll)
- [3] L. Xin, "Source code of our aaai paper on end-to-end target/aspect-based sentiment analysis," 2019. [Online]. Available: <https://github.com/lixin4ever/E2E-TBSA>