

NLU course projects

Edoardo Cecchinato (247434)

University of Trento

edoardo.cecchinato@studenti.unitn.it

1. Introduction

This assignment consists in the implementation of a model, based on a pre-trained version of BERT, for the aspect based sentiment analysis task regarding the extraction of the aspect term only. For this task I took the fine-tuned BERT model defined in assignment 2 and adapt it to the new problem. To test the model I used the `evaluate_ote` function found in the `evals.py` script from the given GitHub repository¹. For the dataset I used the Laptop partition of SemEval2014 task 4 from the same repo.

2. Implementation details

First thing I did for the assignment was to retrieve the data to use to train, validate and test my model. For this I used the dataset from the given repository, specifically the `conll` version since the authors stated that is same data as the original one, but it's already formatted for tasks like name entity recognition. On top of the retrieved data I created `Dataset` and `Dataloader` as I did in the previous assignments.

After this stage I created the model. I started from the BERT model fine-tuned on the tasks of slot filling and intent classification from assignment 2. What I did was removing all the intent related code and adapting the slots part for the new task of aspect extraction, since the two problems are pretty similar. Also in this assignment I used a BERT tokenizer from Hugging Face to convert the text components into tokens and I defined an `untokenize` method to come back to the original words provided these tokens.

For the evaluation part I used the repository `evaluate_ote` function that returns as results the precision, the recall, and the F1 score of the model. What I did was taking the predicted aspects and the ground truth aspects, and convert both of them in their word representation, before passing them to the evaluation function.

3. Results

As I said before the evaluation metrics shown in table 1 are based on the precision, the recall and the F1 score of the model. With all of them I obtained a score of around 0.81, that compared with the assignment 2 BERT results (around 0.95 as F1 score for slots filling task) is not so good. This problem could be related to the fact that in this assignment the dataset size was a little bit smaller than the previous one and the number of aspects much lower than the number of slots. I've also tried to increase and decrease the learning rate, as well as changing the optimizer from Adam to SGD but the results didn't change a lot or were worse. So I kept the configurations shown in table 2. One example of results obtained using SGD as optimizer using a learning rate of 1.0 can be found in table 3. However taking as reference the paper[1] given to us as for the assignment, and looking their results based on the same dataset, it's possible to

see that the scores of my model and their aren't too distance.

| Precision | Recall | F1 score |
|-----------|--------|----------|
| 0.8157 | 0.8170 | 0.8163 |

Table 1: Evaluation results with Adam and $lr=0.00001$.

| Parameter | Value |
|---------------|---------|
| learning rate | 0.00001 |
| optimizer | Adam |
| clip | 5 |
| epochs | 200 |
| patience | 3 |

Table 2: Used parameters.

| Precision | Recall | F1 score |
|-----------|--------|----------|
| 0.8327 | 0.7854 | 0.8083 |

Table 3: Evaluation results with SGD and $lr=0.1$.

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.

¹<https://github.com/lixin4ever/E2E-TBSA>