NLU project exercise lab: 11

Massimo Stefan (240998)

University of Trento

massimo.stefan@studenti.unitn.it

1. Introduction)

In the first part of this project, a two-stage pipeline model was developed to perform Subjectivity and Polarity detection tasks. The pipeline consists of two distinct models: the first model classifies sentences as subjective or objective, while the second model detects the polarity of a document after filtering out objective sentences. Both models were implemented using BERT, a pre-trained language model, and evaluated using a 10-fold cross-validation strategy. The implementation also includes a specific function to filter subjective sentences, contributing to the overall performance of the polarity detection task.

In the second part of this project, the focus shifted to Aspect-Based Sentiment Analysis (ABSA). A joint model was designed to handle both aspects of ABSA, utilizing a pre-trained BERT model for feature extraction and adding specific classifiers for aspect and polarity detection.

2. Implementation details

The implementation of the Part 1 pipeline is orchestrated through two main functions and a main script that seamlessly integrates the entire process. The ‘k_fold_evaluation‘ function is responsible for performing k-fold cross-validation on the given sentences and labels, training and evaluating a BERT model for sequence classification, and ensuring that the distribution of labels is consistent in each fold. The main script structures the overall pipeline by initializing the necessary components such as the tokenizer and loss criterion. It first performs the subjectivity task, then proceeds to the polarity task on the full dataset of movie reviews. Subsequently, the ‘filter_subj_doc‘ function is utilized to filter only the subjective sentences from all the documents in the movie review dataset. Finally, the polarity task is re-run on this filtered dataset, allowing for a comparative evaluation of the results with and without the inclusion of objective sentences.

In Part 2, the focus was on developing a joint model for Aspect-Based Sentiment Analysis (ABSA). The JointABSA class was created, utilizing a pre-trained BERT model for feature extraction. Two separate classifiers were added on top of BERT’s output: the aspect classifier (a token is part of an aspect term) and the polarity classifier (detects the sentiment expressed towards the identified aspect).

In the second phase, the emphasis was on handling the data preparation and alignment. The ABSADataset class was designed to tokenize the sentences and align the annotations with the tokens, considering that a word can be split into multiple tokens. Special handling was implemented to ensure proper alignment and padding, with the Lang class providing mappings between aspect and polarity labels and their corresponding indices.

The JointABSA model was built and trained using the AdamW optimizer, with separate loss functions for aspect term extraction and polarity detection. The Cross-Entropy Loss (CrossEntropyLoss) was used for both tasks, with special handling for padding tokens.

3. Results

The evaluation of the models in part 1 was conducted using a k-fold cross-validation strategy with $k = 10$. This method ensures that the dataset is divided into 10 equal parts, with 9 parts used for training and 1 part for validation in each iteration. The process is repeated 10 times, with each part being used for validation exactly once. The models were evaluated on both the subjectivity and polarity tasks, and the results were compared for the polarity task with and without the removal of objective sentences. This comparative evaluation provided insights into the effectiveness of the filtering strategy implemented in the ‘filter_subj_doc‘ function and its impact on the overall performance of the polarity detection task.

In part 2, a streamlined training and evaluation process was conducted using a custom function (‘train_and_eval‘), targeting the optimization of the joint ABSA model for both aspect term extraction and polarity detection. However, the evaluation procedure encountered an unexpected challenge, as all the metrics (F1, Recall, and Precision for both evaluations) returned a value of 0. This outcome indicates a critical issue in either the model architecture, training process, or evaluation method, necessitating further investigation and refinement to achieve the desired performance in the ABSA task.

Metric	Value
Train loss	0.70
Validation loss	1.39
Accuracy	0.5

Table 1: *Model Performance on the Subjectivity task*

Metric	Value
Train loss	0.69
Validation loss	0.70
Accuracy	0.51

Table 2: *Model Performance on the Subjectivity task*

Metric	Value
Train loss	0.66
Validation loss	0.59
Accuracy	0.6

Table 3: *Model Performance on the Subjectivity task*