NLU project exercise lab: 11

Mateo Rodriguez (239076)

University of Trento

md.rodriguezromero@studenti.unitn.it

For the first part of the lab, there were two tasks at hand, subjectivity and polarity detection. For the first task BERT was used along side some fine-tuning in order to classify objective and subjective sentences, the second model was a simpler SVM that classified if a document is positive or negative. The idea behind the lab is to test the advantageous results from this paper [1] where objective sentence removal improves polarity classification; and indeed there was an increase in performance when objective sentences were removed in the tests performed. For the second part a model was designed to do aspect based sentiment analysis, predicting aspects and their polarity jointly.

1. Introduction

- For the first part of the lab, there are two main models, firts one uses BERT for subjectivity classification, and second one is an SVM was for polarity.
- For the second part the final model used is based on BERT and predicts both aspects and their polarity.

To get a robust estimate of a models' performance the testing was executed using stratified k-folds, with 10 folds. Training for subjectivity was done using the subjectivity dataset and for polarity the movies dataset, both available from NLTK. For the second part the model was trained and tested for 5 runs of 100 epochs each on the laptop dataset from semeval 2014 task 4.

2. Implementation details

Starting with the BERT based model, it uses 2 extra linear layers that process BERT's CLS embeddings and thanks to sigmoid, outputs a value between 0 and 1, where 0 means subjective and 1 objective. Training was done with BCELoss and AdamW. For the polarity task, an SVM with a sigmoid kernel proved to be the best performing. The data was tokenized using TF-IDF and the labels were of course also converted into a binary representation. Now for the combined processing the main problem was the size of the sentences in the documents, because of this it was decided to remove the stopwords reducing the size significantly and not affecting performance. Before removing stopwords the biggest sentence was 15097 characters, and after removing them it drops to 11274 characters, while F1 goes from 0.85 to 0.8498.

For the second part of the lab, since aspect prediction is dependent on the position of words, the data had to be processed to work around BERT's WordPiece tokenizer, which creates subword tokens, severely affecting the model's performance. To do this some regex were created to remove symbols and other characters that led to the creation of sub-word tokens. Following this if a word got subdivided only the first sub-word token is kept and used as input, keeping the tokenized size the same as the text size.

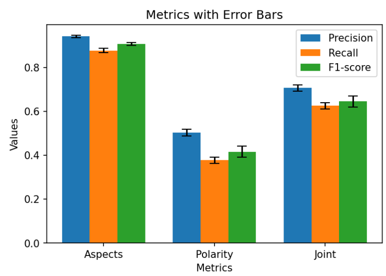
From the BERT backbone the hidden state is processed through three linear layers, with a hidden size of 256. Batch normalization along side the Mish activation function are used to add non-linearity, followed by dropouts of 0.5. The model's output is predicting the polarity of each word in the sentence and by doing so it inherently also contains the mask of which words are aspects, since non-aspect words don't have polarity (at least in the dataset utilized). Nonetheless, testing was done making the model predict both aspects and their polarity separately, but didn't show any improvements compared to the joint prediction. Because of this a single loss is used which is cross entropy.

Initial testing showed the model over-fitting, hindering performance, so augmentations were added to the data. Specifically 2 augmentations were tested [2]. One randomly selects sentences and replaces random words with the most similar words in Word2Vec, and the second augmentation simply chooses a random number of word pairs in a sentence and then proceeds to switch their positions. While testing only the second augmentation was used because Word2Vec proved to be quite slow and acted as a bottleneck for training. The use of the second augmentation helped to delay the over-fitting but in the end more data is needed to increase performance.

3. Results

For the first part of the lab removing objective sentences proved to aid the prediction of polarity, with an average increase in performance of 4.7%. For more details see the annexed figure and table.

For the second part of the lab, results were the following:



For the second part to evaluate aspect prediction and polarity prediction separately a binary mask for the aspects positions was created from the model outputs, which is then compared to the mask of the true positions, while for the polarity evaluation, only the polarity prediction of actual aspects was compared. Finally the joint predictions are the direct outputs of the model. To improve the model's performance more data is needed, specially taking into consideration that more than half of the data has no aspects labeled. Testing reducing the amount of unlabeled examples always showed worse performance.

4. References

[1] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." Ithaca, NY: Cornell University. [Online]. Available: https://aclanthology.org/P04-1035.pdf

[2] D. Raj, "Data augmentation for text," *Analytics Vidhya*, June 2021. [Online]. Available: https://medium.com/analytics-vidhya/data-augmentation-for-text-with-code-6da46aad443d

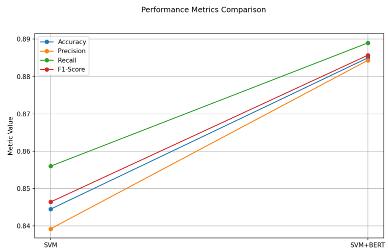


Figure 1: *Polarity prediction with and without objective sentences*

Table 1: *Performance of polarity and subjectivity predictions*

Model	**Accuracy**	**Precision**	**Recall**	**F1-score**
BERT (Subjectivity)	0.99167	0.97952	1.0	0.98838
SVM (Polarity)	0.8445	0.83924	0.856	0.84642
SVM + BERT (Polarity w. subjective sent. only)	0.885	0.88435	0.889	0.8857

Table 2: *Metrics for Aspect-Based Sentiment Analysis*

Model	**Precision**	**Recall**	**F1-score**
Aspect	0.9427 ± 0.0054	0.8779 ± 0.0096	0.9071 ± 0.0065
Polarity	0.5030 ± 0.0154	0.3775 ± 0.0143	0.4163 ± 0.0247
Joint	0.7074 ± 0.0141	0.6256 ± 0.0144	0.6453 ± 0.0248