Vision-Language Pre-Training for Multimodal Aspect-Based Sentiment Analysis

Priyanshu Sharma (862395994), Sahil Chowkekar (862393156)

March 18, 2023

1 General Descriptions

- The complete code is present in this **github/priyanshu-sharma/VLM**. Presently, it is in a development state, so we keep it in a private repo. But anyone can request access to this repo.
- We use two different datasets, as stated in the original paper, for downstream task. These datasets are downloadable from this **link** and collectively known as **MVSA-Multi** and comprises of "Twitter - 2015" and "Twitter - 2017" datasets.
- For our final proposed method, we added the three different type of pooling technique to our BART-based approach and we study the impact of the same our final model. For comparison, we also evaluated the BERT-based approaches such as TomBERT, and mBERT with different pooling types.
- We consider all the three subtasks in MABSA as our downstream tasks, including Joint Multimodal Aspect-Sentiment Analysis (JMASA), Multimodal Aspect Term Extraction (MATE), and Multimodal Aspect-oriented Sentiment Classification (MASC) along with all the different configuration of Multimodal BERT-based approach with different pooling techniques.
- We also evaluate the multimodal aspect-based sentiment classification using the BERT-based approach along with the BART-based one. And we found that BART-based approaches along with the different pooling can outperform the previous BART-based approaches such as mBERT and TomBERT with base BERT model. We extensively tested this approach and measure the performance with the help of MicroF1 score (F1), Precision (P) and Recall (R) as the evaluation metrics. Overall, for downstream tasks, we presented the results using two sections:-
  - **BART Evaluation Metrics** - We have compared the five different evaluation metrics for all three pooling techniques, which are: - Dev Recall, Dev MicroF1 Score, Loss, Dev Precision and MASC Accuracy.

  - **BERT Evaluation Metrics** - We have compared the three different multimodal BERT-based approaches such as TomBERT, mBERT, and ResBERT along with two different pooling technique to obtain the final output.

- Other than evaluating the performance with Base Model we also evaluated the performance with Large Models. For BART-based approaches, we observe that as increase the size of the pre-trained model our accuracy on downstream tasks keeps on increasing, while, with BERT-based approaches, we didn't observe such a case.
- For the BERT-based approach and dev set, we get an accuracy of 68.37%, and micro F1 of 64.84% and with the test set, we get an accuracy of 67.50%, a precision of 66.29%, recall of 65.29% and microF1 of 65.75%.
- For the BART-based approach and dev set, we get an accuracy of 75.04%, a recall of 69.86%, and a precision of 69.38% and with the test set, we get an accuracy of 76.18% and a precision of 71.66%.
- Contact the authors of this report for more information.

2 BART Evaluation Metrics

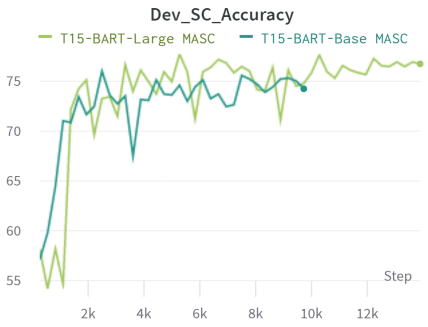

Figure 1: Dev Accuracy

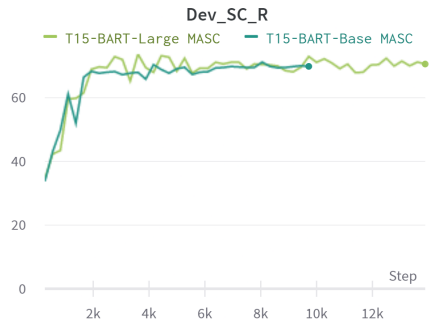

Figure 2: Dev Recall

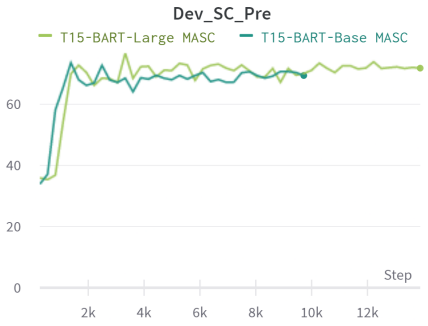

Figure 3: Dev Precision

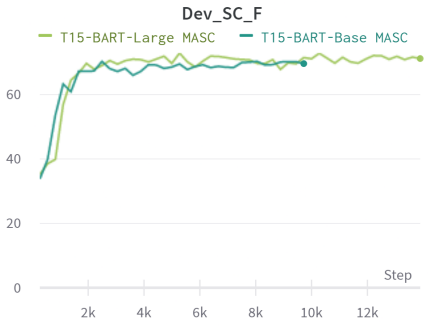

Figure 4: Dev MicroF1

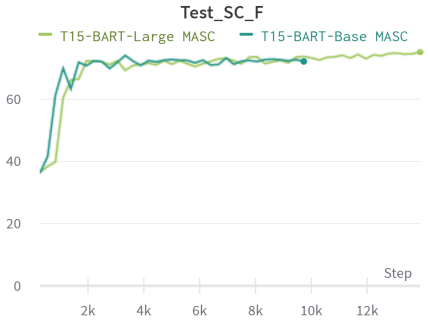

Figure 5: Test MicroF1

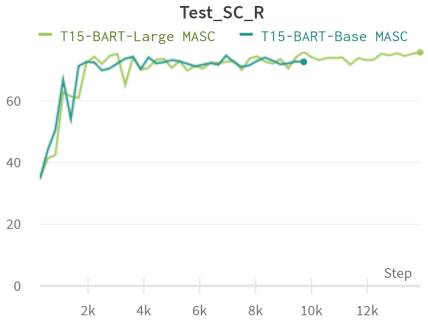

Figure 6: Test Recall

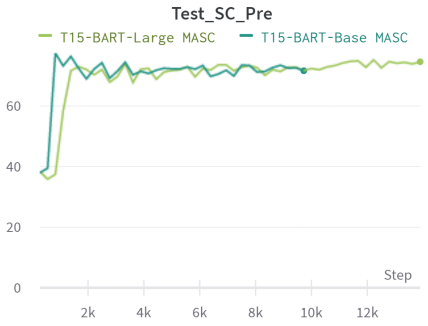

Figure 7: Test Precision

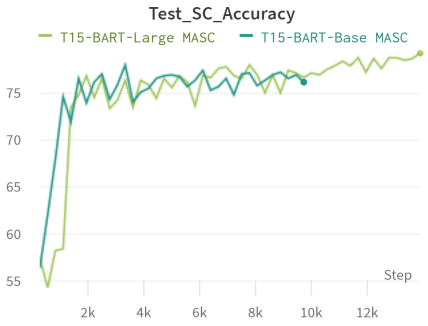

Figure 8: Test Accuracy

3 BERT Evaluation Metrics

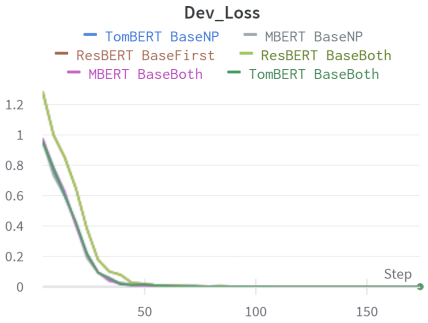

Figure 9: BERT Dev Loss

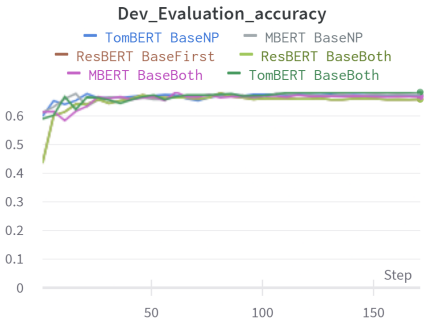

Figure 10: BERT Dev Evaluation Accuracy

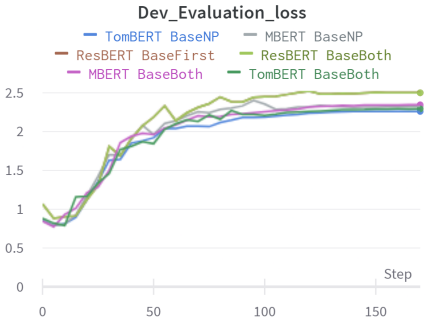

Figure 11: BERT Dev Evaluation Loss

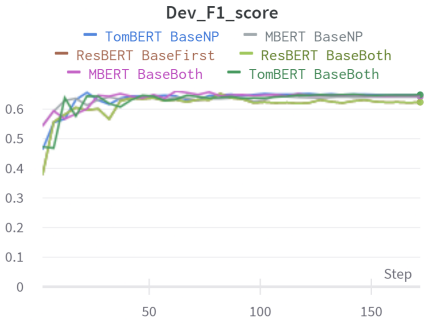

Figure 12: BERT Dev F1 Score

4 Results

- BART-based approaches clearly outperform the BERT-based approaches along with different pooling technique on both the datasets.
- Best Performance is achieved by the Large-BART model with "BOTH" pooling technique, which is around 79.29% accuracy, 74.63% precision, recall of 75.86% and F1 score of 75.24% on Test Dataset.
- Best Performance for BERT-based approaches is achieved by the TomBERT with "BOTH" pooling technique, which is around 69.04% accuracy, precision of 67.78%, recall of 66.08% and F1 score of 66.83% with test loss of 0.00025 on Test Dataset. In general, mBERT performs better than ResBert and BOTH and FIRST pooling technique is better than CLS pooling technique.
- We also evaluated the performance of BERT-based approaches with a large BERT model but found that it degrade the performance significantly.
- Overall, we observed the higher memory and GPU utilization for larger model across all the different tasks.