

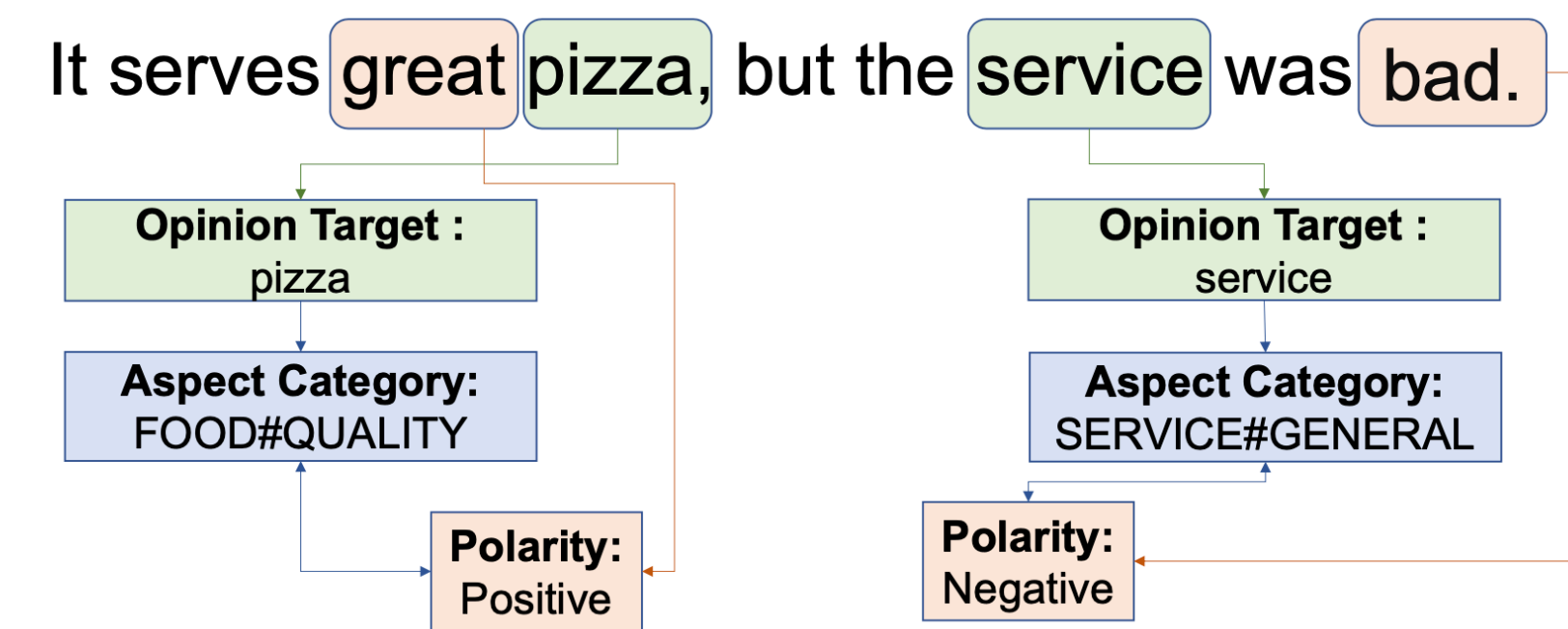
ABSA-Bench: Towards the Unified Evaluation of Aspect-based Sentiment Analysis Research

Abhishek Das
 School of Computer Science
 University of Adelaide
abhishek.das@adelaide.edu.au

Wei Emma Zhang
 School of Computer Science
 University of Adelaide
wei.e.zhang@adelaide.edu.au

Background

- ABSA is the task of identifying fine-grained opinion polarity towards a specific aspect associated with a given target.



- Recent progress have been made with the advancement of Deep Neural Network.
- Attention mechanism has played an important role outperforming previous approaches by paying more attention to the context words that are semantically-closer with the aspect terms.
- The most recent approaches adopted pre-trained Bidirectional Encoder Representations from Transformers (BERT) architecture.

Motivation

- Lack of unbiased comparison overall.
- Existing benchmarking research works are mostly conducted on evaluating single tasks and none of them support aspect-based sentiment analysis.
- We fill this gap by proposing a unified evaluation process and building a united platform for comparing different ABSA models.

ABSA-Bench

- A Web-based benchmark platform that enables researchers to evaluate their ABSA models.
- The platform consists of three primary elements: Leader board, Evaluation Portal, and Discussion forum.
- Leader Board :The performances of the models that are submitted by the authors will be added to the leader board and assigned a proper ranking.
- Evaluation Portal : Consist of Jupyter-Hub that is served by the computational power of the Google Cloud Platform. Users can train and evaluate their model in their own workspaces leveraging the resources.
- Discussion forum : Serve as a collaborative environment where researchers can post queries and collaborate for future research.

Results

Models	Restaurant	Laptop
CNN	60.25	57.75
LSTM	65.51	55.35
TD-LSTM	68.98	61.87
TC-LSTM	66.72	61.11
ATAE-LSTM	63.72	58.47
CABASC	68.02	62.94
IAN	65.12	60.90
RAM	66.76	59.73
MemNet	61.09	58.01
AEN-BERT	73.76	76.31
BERT-PT	76.96	75.08
BERT-SPC	73.03	72.63
LCF-BERT	81.74	79.59

Table1: Performance score of different model in terms of Macro F1. LCF-BERT model provided the best performance among BERT based models while CABASC has obtained the highest F1 score on both datasets for Non-BERT based models.

Models	Params 10 ⁶	Memory (MB)
CNN	1.21	10.01
LSTM	7.23	35.61
TD-LSTM	1.44	12.41
TC-LSTM	2.16	14.11
ATAE-LSTM	2.53	16.61
CABASC	1.53	12.61
IAN	2.16	16.18
RAM	6.13	31.18
MemNet	0.36	7.82
AEN-BERT	112.93	451.84
BERT-PT	110	450.23
BERT-SPC	109.48	450.58
LCF-BERT	113.61	452.62

Table 2: Comparison of different model size in terms of parameters and memory. BERT based models have are much bigger in comparison with Non-BERT based models.

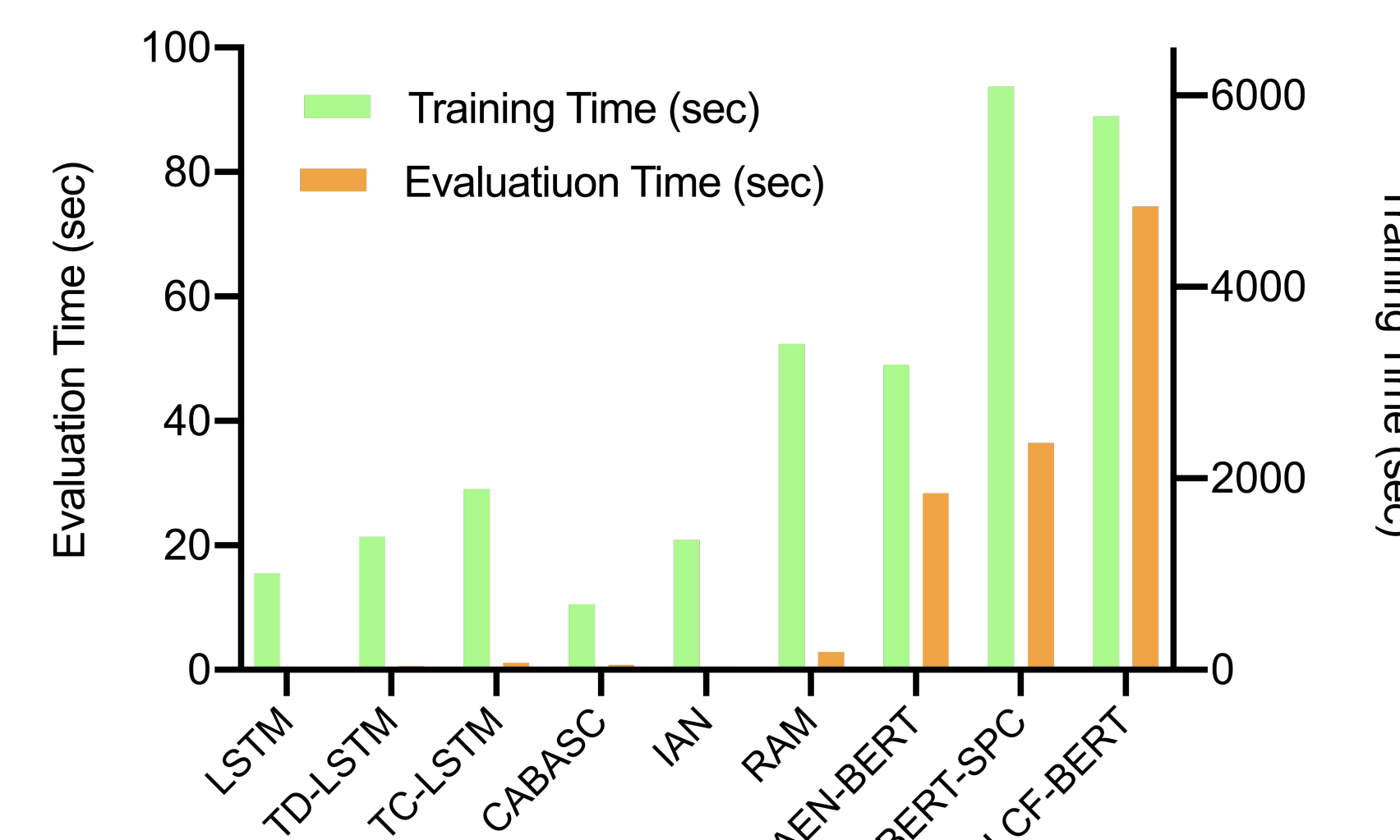


Table 3: While BERT based models overall performed much better than Non-BERT based models, it is computationally more expensive.

Conclusions and Future Directions

- Design and implemented an ABSA benchmarking evaluation platform which will help to understand the implementation of different deep learning models performing the task of ABSA that can be further utilised to improve the existing models.
- Updating our benchmarking platform with new tasks and datasets which will encourage quantitatively informed research and learning.