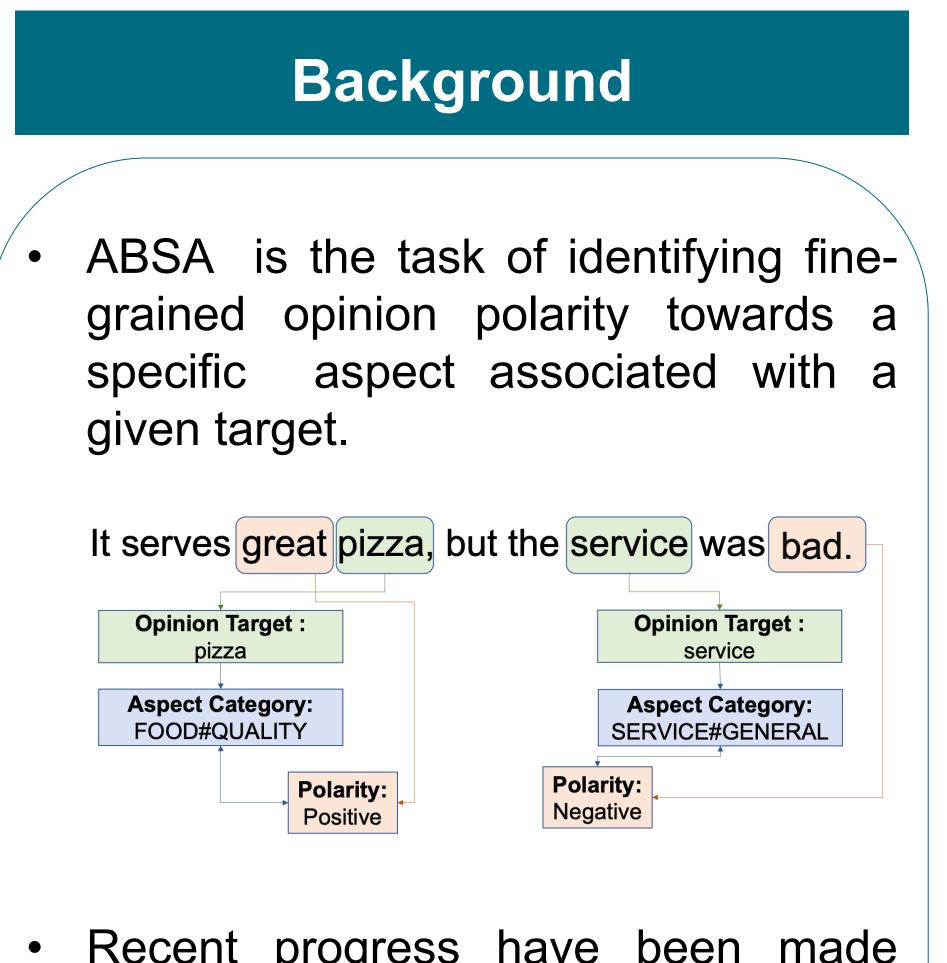
ABSA-Bench: Towards the Unified Evaluation of Aspect-based Sentiment Analysis Research



- 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 context words that are the to 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 \bullet 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 lacksquareevaluation process and building a united platform for comparing different ABSA models.

Abhishek Das

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

					Home	Leaderboard	Evaluation Portal	Discussions A	ccount
					Ber	nchmarking A-Bench is a ben p learning appro	-BEN state-of-the-a	rt in ABSA for evaluating and d	
					Home	Leaderboard	Evaluation Port	al Discussions	Account
BEI	RT Leac		rd						
Ra	ink Model	Restaurant	Laptop	Paper					
Rai 1	LCF-	Restaurant 81.74	Laptop 79.59	Paper LCF: A Local Context Focus M	1echanism	for Aspect-Bas	sed Sentiment Clas	ssification.	
Rai 1 2									
1	LCF- BERT	81.74	79.59	LCF: A Local Context Focus M Bert post-training for review	reading co	omprehension	and aspect-based		
1	LCF- BERT BERT-PT AEN-	81.74 76.96	79.59 75.08	LCF: A Local Context Focus M Bert post-training for review analysis.	reading co	omprehension	and aspect-based		
1	LCF- BERT BERT-PT AEN-	81.74 76.96	79.59 75.08	LCF: A Local Context Focus M Bert post-training for review analysis.	reading co	omprehension	and aspect-based		
1	LCF- BERT BERT-PT AEN-	81.74 76.96	79.59 75.08	LCF: A Local Context Focus M Bert post-training for review analysis.	reading co	omprehension	and aspect-based	l sentiment	Account
1 2 3	LCF- BERT BERT-PT AEN-	81.74 76.96	79.59 75.08	LCF: A Local Context Focus M Bert post-training for review analysis.	reading co	omprehension (and aspect-based	l sentiment	Account
1 2 3 FO	LCF- BERT BERT-PT AEN- BERT	81.74 76.96 73.76	79.59 75.08	LCF: A Local Context Focus M Bert post-training for review analysis. Attentional Encoder Network	reading co	omprehension (and aspect-based	l sentiment	Account
1 2 3 FO F	LCF- BERT BERT-PT AEN- BERT	81.74 76.96 73.76	79.59 75.08 76.31	LCF: A Local Context Focus M Bert post-training for review analysis. Attentional Encoder Network	reading co	omprehension (and aspect-based	I sentiment	Account
1 2 3 3 FO FO FO This	LCF- BERT BERT-PT AEN- BERT	81.74 76.96 73.76 	79.59 75.08 76.31	LCF: A Local Context Focus M Bert post-training for review analysis. Attentional Encoder Network	reading co	omprehension (and aspect-based	I sentiment	Account
1 2 3 3 FO FO FO This I RSS	LCF- BERT BERT-PT AEN- BERT	81.74 76.96 73.76 	79.59 75.08 76.31	LCF: A Local Context Focus M Bert post-training for review analysis. Attentional Encoder Network	reading co	omprehension (and aspect-based	I sentiment	Account
1 2 3 3 FO FO FI I SS DIS DIS	LCF- BERT BERT-PT AEN- BERT SCUUSSIONS	81.74 76.96 73.76 73.76 Re egory / section	79.59 75.08 76.31	LCF: A Local Context Focus M Bert post-training for review analysis. Attentional Encoder Network	reading co	omprehension (and aspect-based	I sentiment	Account
1 2 3 3 FO FO FO This I RSS Dis U I RSS	LCF- BERT BERT-PT AEN- BERT SCUSSIONS is a simple cate Sacussions	81.74 76.96 73.76 73.76 Re egory / section	79.59 75.08 76.31	LCF: A Local Context Focus M Bert post-training for review analysis. Attentional Encoder Network	reading co	omprehension of ed Sentiment C	and aspect-based	I sentiment	

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.

Results

Models	Params 10 ⁶	Memory					
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.

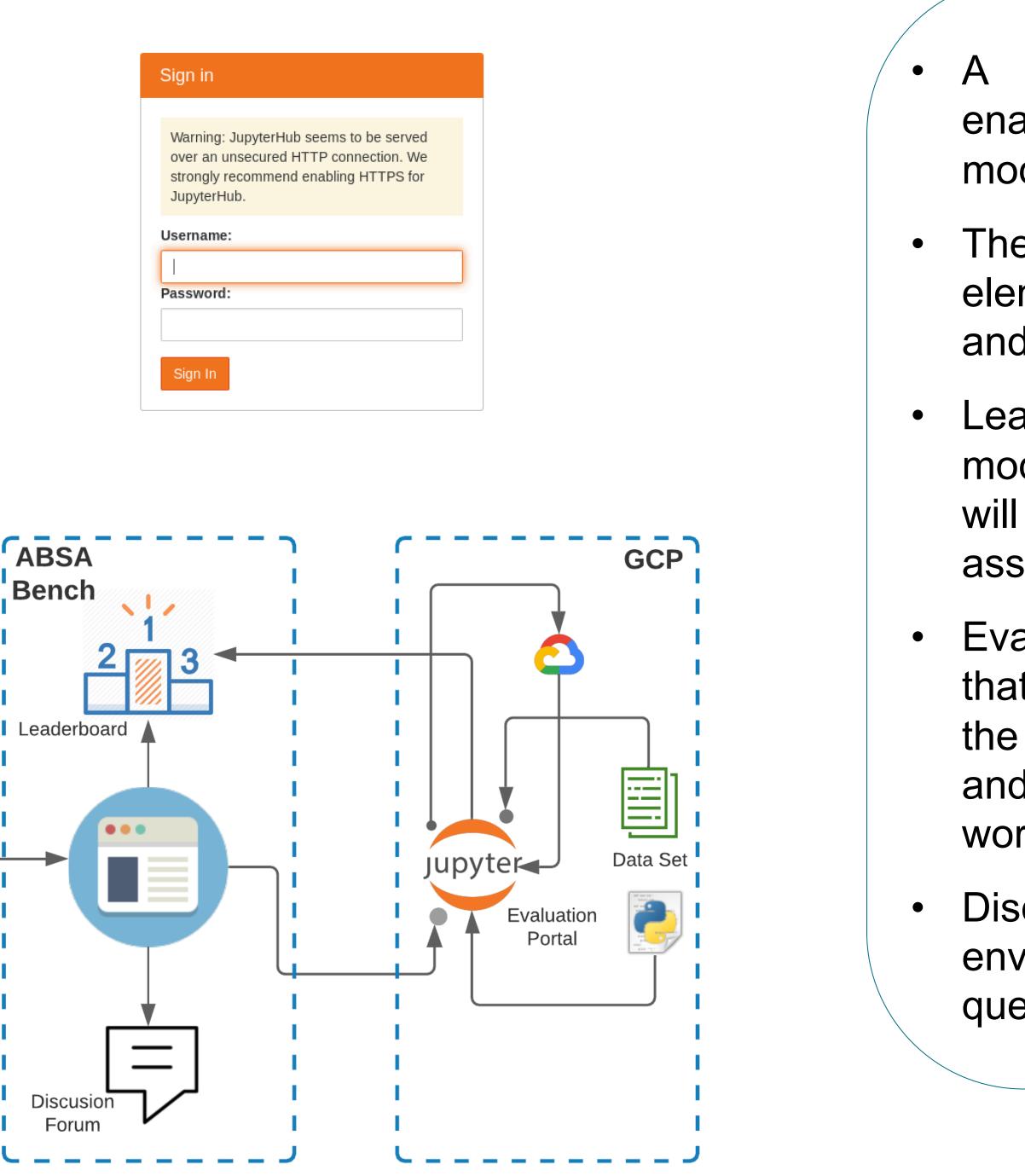
Acknowledgments: This project is sponsored by Google Academic Research Grants.

Wei Emma Zhang

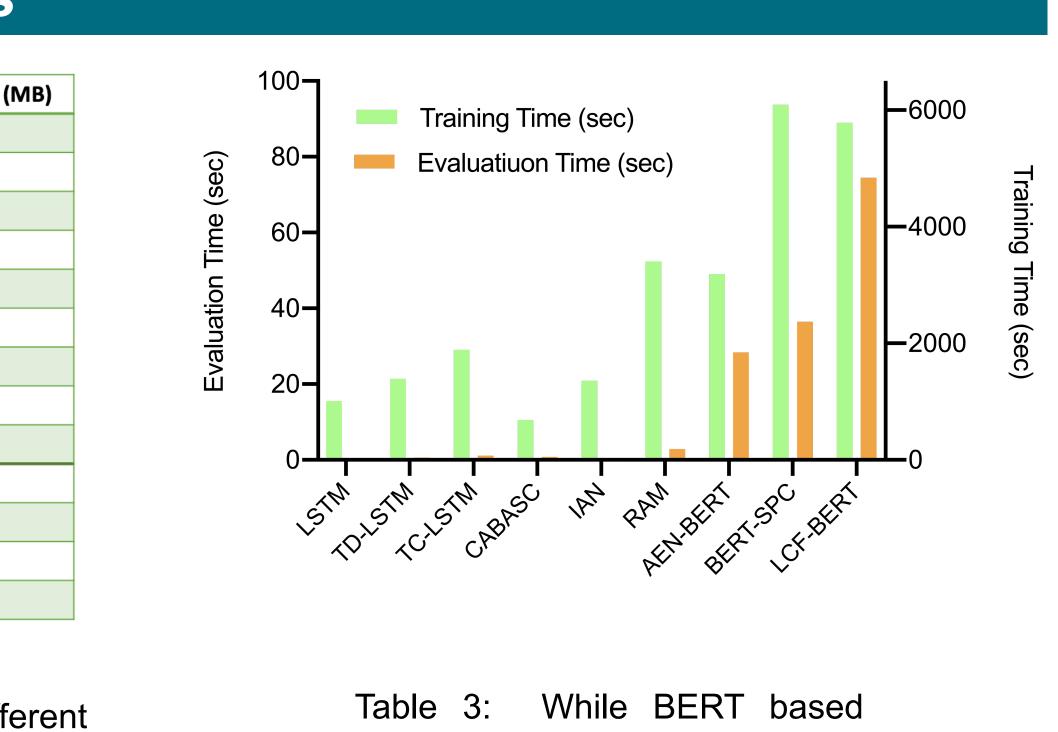
School of Computer Science University of Adelaide wei.e.zhang@adelaide.edu.au

ABSA-Bench

Jupyter







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

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.

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.