Aspect-specific Context Modeling for Aspect-based Sentiment Analysis

Fang Ma, Chen Zhang, Bo Zhang, Dawei Song
Beijing Institute of Technology Beijing
{mfang, czhang, bo.zhang, dwsong}@bit.edu.cn

北京理工大學
BEIJING INSTITUTE OF TECHNOLOGY

Outline

- Background: Aspect-based Sentiment Analysis (ABSA)
- Motivation
- Method: Aspect-specific input transformations
- Experiment
- Results and Analysis
- Conclusion

Background

- Aspect-based Sentiment Analysis (ABSA)
  - Fine-grained opinion mining
  - May contains multiple aspects
- Two sub-tasks of ABSA
  - Aspect-based Sentiment Classification (SC)
  - Aspect-based Opinion Extraction (OE)
    - E.g., The *food is tasty but the service is very bad !*

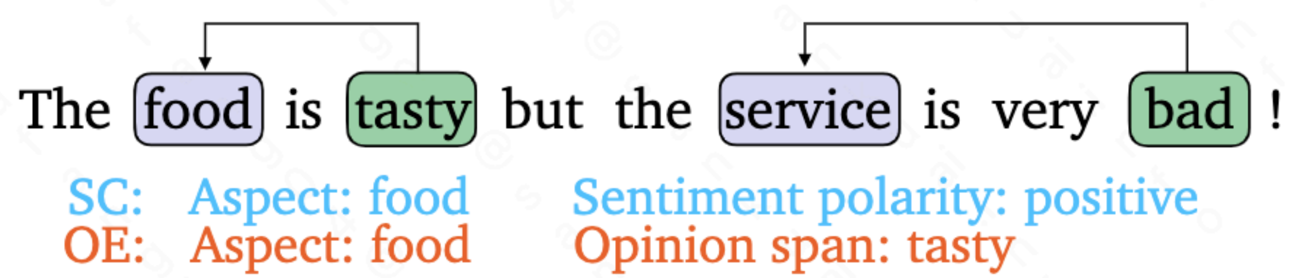


Figure 1: Example of the SC and OE. The words highlighted in purple represent the given aspects, whereas the words in green represent the corresponding opinion.

Motivation

- An effective ABSA model requires either *aspect-specific feature induction* or *context modeling*.
- Prior ABSA work relies on rather complicated aspect-specific feature induction to achieve a good performance.
  - E.g., for SC, these approaches range from memory networks, convolutional networks, attentional and graph-based networks.
  - E.g., OE is treated as a sequence tagging task, attention-based and graph-based networks are developed to capture the interaction between the aspect and the context.
- Recently, pre-trained language models (PLMs) have been shown to enhance the sofa ABSA models due to their extraordinary context modeling ability.

Motivation

- However, currently the use of PLMs in ABSA models is *aspect-general*, which only use the PLM as context modeling layers to simplify the feature induction strictures, overlooks two key questions:
  - Whether the context modeling of a PLM can be *aspect-specific*?
  - Whether the *aspect-specific context modeling* within a PLM can further enhance ABSA?
- To answer the questions, we propose to achieve *aspect-specific context modeling* with *aspect-specific input transformations*, namely *aspect companion*, *aspect prompt*, and *aspect marker*.
- Informed by these transformations, non-intrusive aspect-specific PLMs can achieved to promote the PLM to pay more attention to the aspect-specific context in a sentence.
- Additionally, we craft an adversarial benchmark for ABSA (advABSA) to see how aspect-specific modeling can impact model robustness.

3.Method

- 3.1 Task Description
  - Sentence $S = \{w_1, w_2, \ldots, w_n\}$
  - Given Aspect $A = \{a_1, a_2, \ldots, a_m\}$
  - The goal of SC is to find the sentiment polarity with respect to the given aspect A.
    - SC requires a model to give a positive sentiment on *food*
  - OE aims to extract corresponding opinion span based on the given aspect A.
    - OE requires a model to tag the sentence as {O, O, O, B, O, O, O, O, O, O, O,}, indicating the opinion span *tasty* for the aspect *food*.

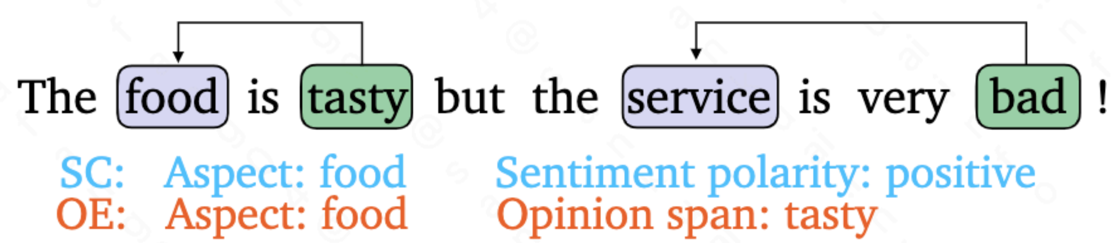


Figure 1: Example of the SC and OE. The words highlighted in purple represent the given aspects, whereas the words in green represent the corresponding opinion.

Method

- 3.2 Overall Framework
  - An input transformation layer
    - *Aspect Generality*
    - *Aspect Companion*
    - *Aspect Prompt*
    - *Aspect Marker*
  - A context modeling layer
  - A feature induction layer
  - A classification layer

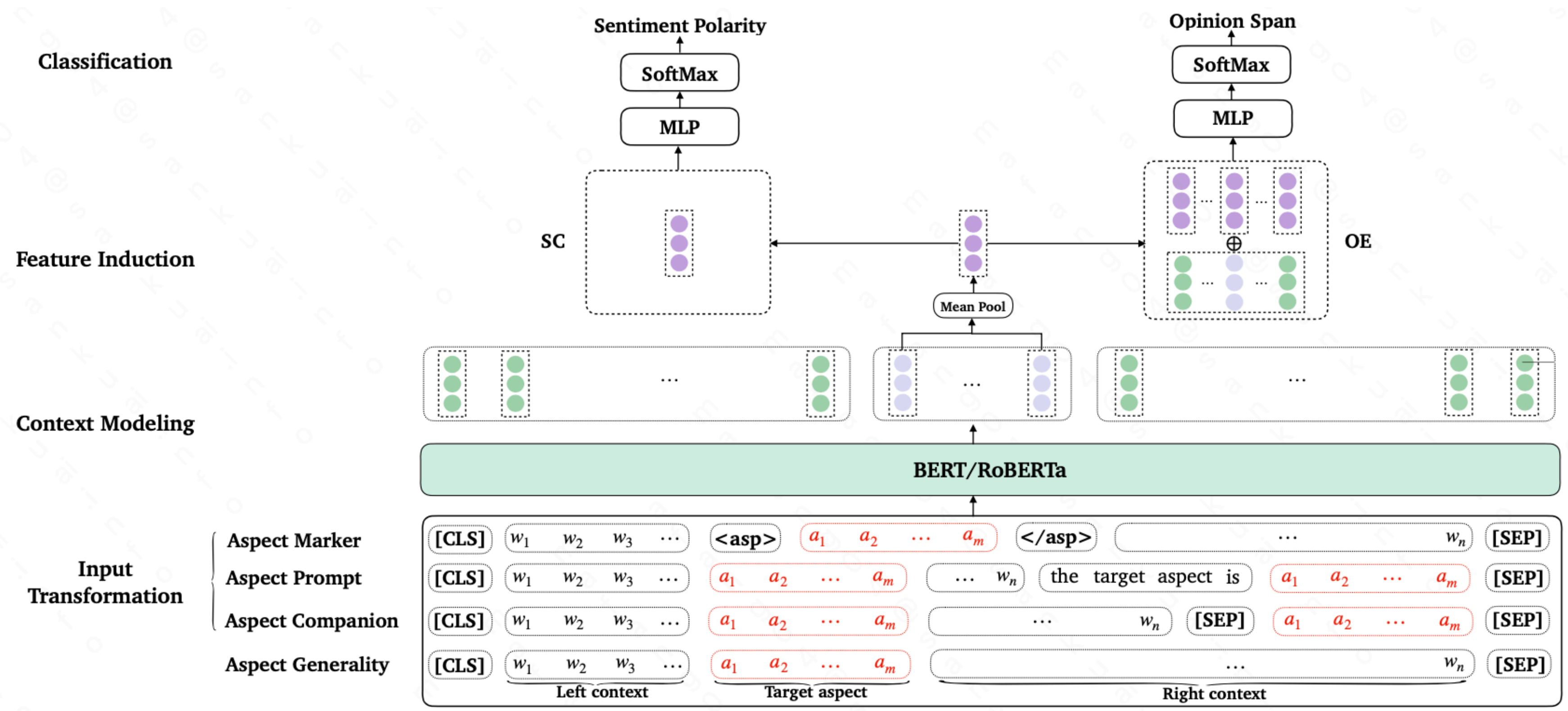
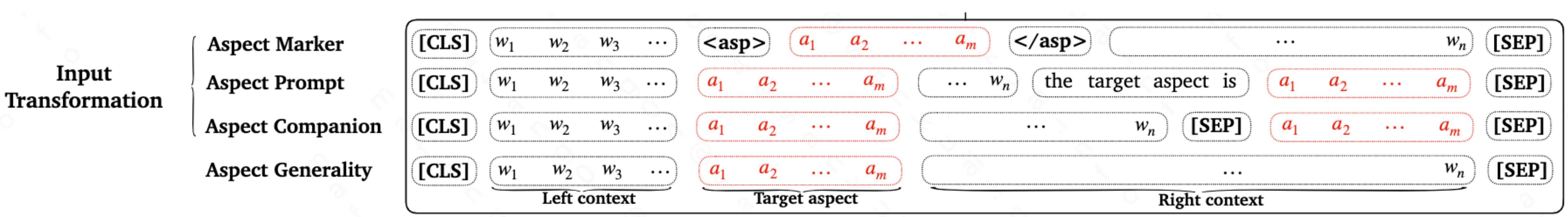


Figure 2: The architecture of our proposed model based on the three mechanisms.

Method

- 3.3 Aspect-general Input
  - Aspect Generality
    - $[CLS]\{w_1, w_2, \ldots, \}\{a_1, a_2, \ldots, a_m,\}\{\ldots, w_n\}[SEP]$
- 3.4 Aspect-specific Input Transformation
  - We hypothesize that the three transformations can promote the aspect-awareness of PLM and help PLM achieve an effective aspect-specific context modeling.
  - 3.4.1 *Aspect Companion*
    - $[CLS]\{w_1, w_2, \ldots, \}\{a_1, a_2, \ldots, a_m,\}\{\ldots, w_n\}[SEP]\{a_1, a_2, \ldots, a_m,\}[SEP]$
  - 3.4.2 *Aspect Prompt*
    - $[CLS]\{w_1, w_2, \ldots, \}\{a_1, a_2, \ldots, a_m,\}\{\ldots, w_n\}$ *the target aspect is* $\{a_1, a_2, \ldots, a_m,\}[SEP]$
  - 3.4.3 *Aspect Marker*
    - $[CLS]\{w_1, w_2, \ldots, \}\ \langle asp \rangle\ \{a_1, a_2, \ldots, a_m,\}\ \langle /asp \rangle\ \{\ldots, w_n\}[SEP]$



Method

- 3.5 Context Modeling

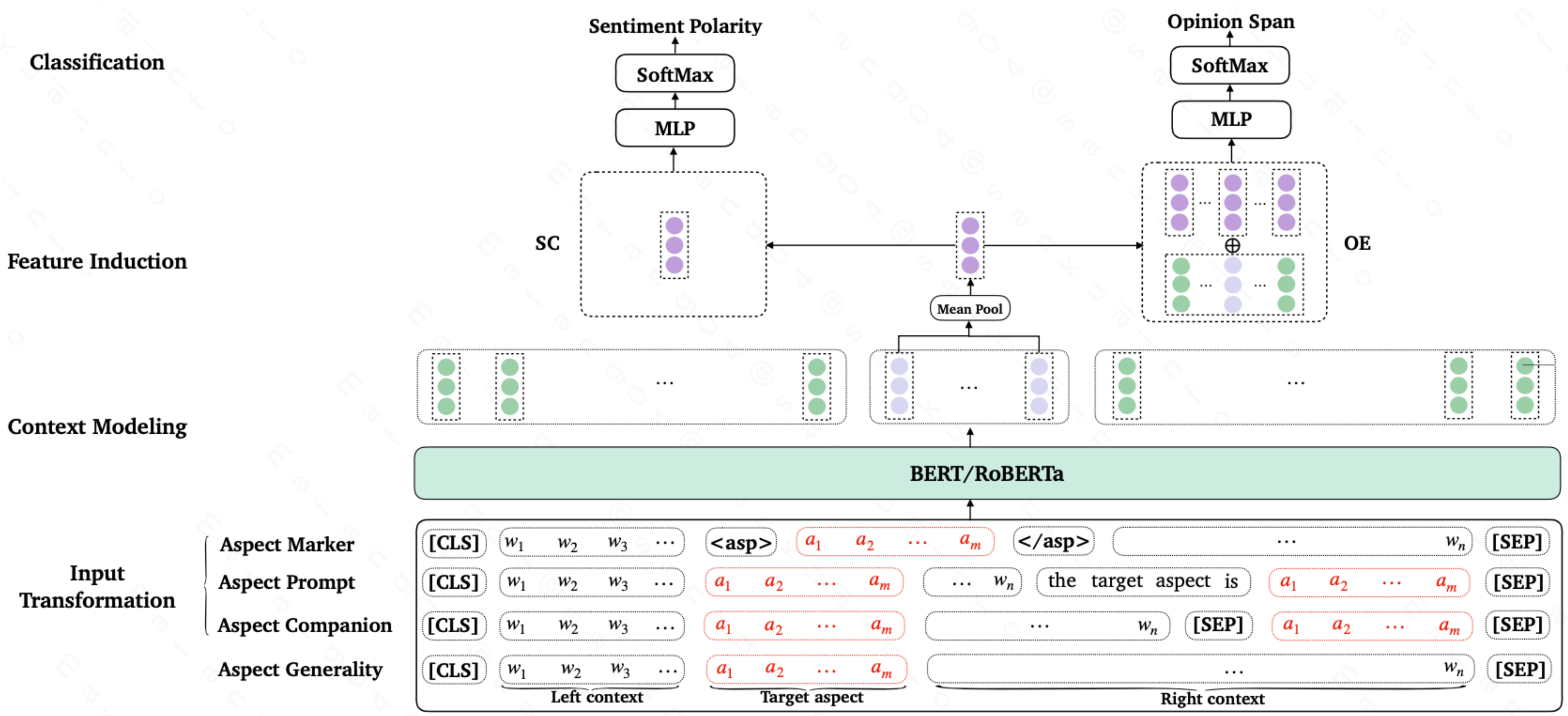
$$H = \mathrm{PLM}(\hat{S})$$

- 3.6 Feature Induction

$$\hat{H} = \mathrm{MeanPool}([h_1^a, h_m^a])$$

- 3.7 Fine Tune

$$\mathscr{L}(\theta) = -\sum_{i=1}^{n} \hat{y}_i \log y_i + \lambda \sum_{\theta \in \Theta} \theta^2$$



Experiment

- SC Datasets
  - Standard datasets
    - SemEval Restaurant
    - SemEval Laptop
  - Robustness datasets
    - ARTS-SC-Restaurant
    - ARTS-SC-Laptop
- SC Baselines (PLM-based)
  - Bert/Roberta-CLS-MLP
  - AEN-Bert
  - LCF-Bert
  - PLM-ASCNN
  - PLM-ASGCN
- OE Datasets
  - Standard datasets
    - SemEval Laptop
    - SemEval Restaurant
  - Robustness datasets
    - ARTS-OE-Laptop
    - ARTS-OE-Restaurant
- OE Baselines (PLM-based)
  - Bert+Distance-rule
  - TF-Bert
  - SDRN
  - TSMSA-Bert
  - ARGCN-Bert

Generation Strategy	Target Aspect: Opinion	Other Aspepct:Opinion	Example
Source: The original sample from the test set	works : well positive	apple OS : happy	Works well , and I am extremely happy to be back to an apple OS .
RevTgt: Reverse the sentiment of the target aspect	works : badly negtive	apple OS : happy	Works badly , but I am extremely happy to be back to an apple OS .
RevNon: Reverse the sentiment of the non-target aspects with originally the same sentiment as target	works : well positive	apple OS : unhappy	Works well , but I am extremely happy to be back to an apple OS .
AddDiff: Add aspects with the opposite sentiment from the target aspect	works : well positive	apple OS : happy games : issue video chat : iffy	Works well , and I am extremely happy to be back to an apple OS , but games being the main issue . And the video chat is the only thing that is iffy about it .

Table 7: The example of using three adversarial strategies to generate the Aspect Robustness Test Set with spans (**ARTS-OE**) based on SemEval. Specifically, we use these strategies to generate 1002 test instances for the laptop domain (**ARTS-OE-Lap**) and 2009 test instances for the restaurant domain (**ARTS-OE-Res**). Each aspect in a sentence is associated with an opinion span for OE.

Results and Analysis

- SC Standard and Robustness Results
  - Our models with input transformations outperform the comparative baseline models
  - Our models with the three transformations are more robust than the baseline models

Models	SEM-LAP				SEM-REST			
	Standard		Robustness		Standard		Robustness	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
IAN	67.74	59.99	52.91	47.54	77.48	66.39	57.75	48.12
Memnet	67.81	60.67	52.00	46.50	76.77	64.46	55.30	46.67
AOA	69.47	63.13	52.00	46.50	77.57	66.02	58.19	49.02
ASGCN	70.97	65.31	56.59	52.12	78.87	68.12	64.89	55.41
AEN-BERT	77.37	71.83	71.49	66.37	83.66	75.50	73.24	66.31
LCF-BERT	76.55	71.40	71.19	66.95	81.66	72.24	70.57	62.75
BERT-CLS+MLP	75.42	69.08	54.91	51.21	78.95	67.66	53.86	47.16
RoBERTa-CLS+MLP	79.09	75.36	56.24	54.61	81.93	71.19	60.45	52.02
BERT-ASCNN	76.33	71.09	71.17	66.90	82.66	74.05	75.73	68.17
RoBERTa-ASCNN	81.41	77.22	73.59	70.14	85.93	78.01	78.85	70.69
RoBERTa-ASGCN	81.82	78.28	73.48	69.38	85.66	78.48	79.65	72.56
BERT-MeanPool	76.87	71.71	70.59	66.38	84.27	76.48	77.36	70.64
+AC	75.30	69.62	69.40	64.45	84.12	76.16	76.78	69.86
+AP	76.39	70.91	68.92	63.77	83.89	76.02	76.48	69.34
+AM	76.33	**71.93**$\uparrow$0.22	**70.78**	**67.06**$\uparrow$0.68	**84.71**	**78.07** $\uparrow$1.59	**78.10**	**72.38** $\uparrow$1.74
RoBERTa-MeanPool	81.38	77.68	74.67	71.21	85.41	78.15	79.75	72.73
+AC	**81.54**	77.54	**75.13**	71.02	**86.68**†	**79.69**†$\uparrow$1.54	**80.63**	**74.03**$\uparrow$1.30
+AP	**81.85**	**77.91**$\uparrow$0.23	74.53	70.48	**86.43**	**79.43**$\uparrow$1.28	**80.72**	**74.09**†$\uparrow$1.36
+AM	**82.07**†	**78.50**†$\uparrow$0.82	**75.90**†	**72.59**†$\uparrow$1.38	**86.41**	**79.58**$\uparrow$1.43	**80.88**†	**74.04**$\uparrow$1.31

Table 3: Standard and robust experimental results (%) on SC. The first and second blocks indicate non-PLM and PLM-based baseline models. Our models and better results are bold (Acc and F1, the larger, the better). The marker † represents that our models outperform the all other models significantly ($p < 0.01$), and the small number next to each score indicates performance improvement ($\uparrow$) compared with our aspect-general base model (BERT-MeanPool/RoBERTa-MeanPool).

Results and Analysis

- OE Standard and Robustness Results
  - With the three transformations, our models perform significantly better than baseline models.
  - With the transformations, our models are more robust.

Models	SEM-LAP		SEM-REST	
	Standard	Robustness	Standard	Robustness
Pipeline*	63.83	-	69.18	-
IOG*	70.99	-	80.23	-
LOTN*	72.02	-	82.21	-
ARGCN*	75.32	-	84.65	-
BERT+Distance-rule*	70.54	-	76.23	-
TF-BERT*	72.26	-	78.23	-
SDRN*	80.24	-	83.53	-
TSMSA-BERT*	82.18	-	86.37	-
ARGCN-BERT*	76.36	-	85.42	-
BERT-MeanPool-Concat	68.27	39.68	69.08	44.23
+AC	80.31↑**12.04**	70.98↑**31.30**	85.09↑**16.01**	70.01↑**25.78**
+AP	79.60↑**11.33**	68.06↑**28.38**	85.32↑**16.24**	70.25↑**26.02**
+AM	81.06 ↑**12.79**	71.23↑**31.55**	85.62↑**16.54**	69.68↑**25.45**
RoBERTa-MeanPool-Concat	69.74	38.76	79.03	56.93
+AC	82.78↑**13.04**	71.26↑**32.50**	86.03↑**7.00**	71.42↑**14.49**
+AP	82.63↑**12.89**	71.46↑**32.30**	**86.58**† ↑**7.55**	**71.61**† ↑**14.68**
+AM	**83.83**† ↑**14.09**	**73.69**† ↑**34.93**	86.33↑**7.30**	71.50↑**14.57**

Table 4: Standard and robustness evaluation results (F1-score, %) on OE. The first and second blocks show the results of the non-PLM and BERT-based baseline models (with $*$) respectively, which are extracted from the published papers (Wu et al., 2020) and (Feng et al., 2021). Note that there were no robustness results of the baseline models in the original published papers, so that we leave then blank. The results of our models are presented in the third and fourth blocks. The best results are bold (F1-score, the larger, the better).

Ablation Study

- Aspect-specific Context Modeling
  - The results show that the transformations bring significant performance improvements, even better than the models with aspect feature induction.
  - These excellent results demonstrate the effectiveness of the proposed transformations for context modeling, which indirectly explains that context modeling is more critical than aspect feature induction for ABSA.
- Aspect-specific Feature Induction
  - These results demonstrate the effectiveness of the aspect-specific feature induction methods with PLMs.

Models	SEM-LAP	SEM-REST
BERT-MeanPool	71.71	76.48
BERT-CLS+MLP	69.08	67.66
+AC	68.82	74.03↑**6.37**
+AP	70.47↑**1.39**	76.78↑**9.12**
+AM	70.24↑**1.16**	74.19↑**6.53**
RoBERTa-MeanPool	77.68	78.15
RoBERTa-CLS+MLP	75.36	71.19
+AC	77.62↑**2.26**	76.04↑**4.85**
+AP	78.40↑**3.04**	78.53↑**7.34**
+AM	78.21↑**2.85**	79.91↑**8.72**

Table 5: SC ablation experimental results (F1-score, %).

Models	SEM-LAP	SEM-REST
BERT-MeanPool-Concat	68.27	69.08
BERT-MLP	67.67	61.40
+AC	79.95↑**12.28**	79.46↑**18.06**
+AP	80.08↑**12.41**	81.02↑**19.62**
+AM	81.50↑**13.83**	80.02↑**18.62**
RoBERTa-MeanPool-Concat	69.74	79.03
RoBERTa-MLP	67.92	60.00
+AC	82.18↑**14.26**	81.59↑**21.59**
+AP	81.96↑**14.04**	81.04 ↑**21.04**
+AM	83.42↑**15.50**	80.81↑**20.81**

Table 6: OE ablation experimental results (F1-score).

Visualization of Attention

- The attention scores separately offered by our OE model (BERT-MeanPool-Concat) with the three transformations
  - We can observe that before applying the transformations, the model may attend to more irrelevant words.
  - On the contrary, AC, AP, and AM can promote our model to attend to aspect-specific context and capture the correct opinion spans, thus achieving aspect-specific context modeling in PLM.

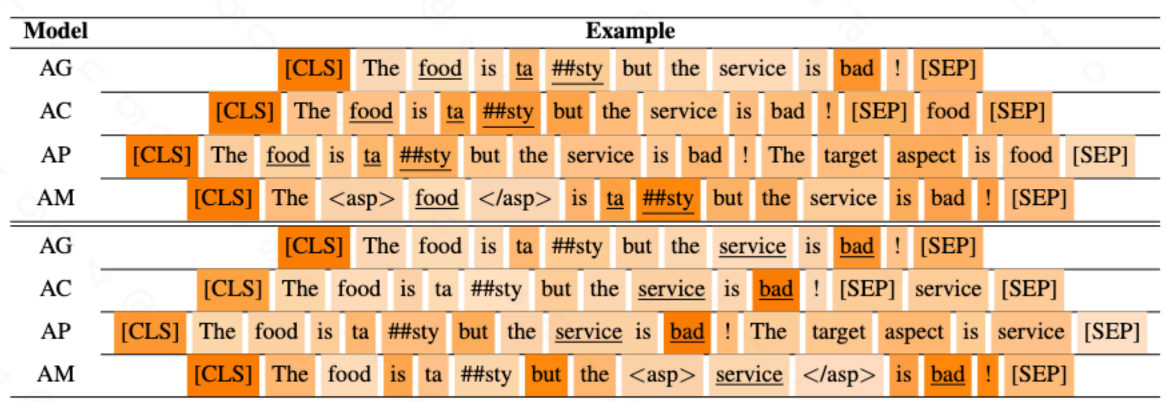


Figure 3: **Attention visualization**. Gradient saliency maps (Simonyan et al., 2014) for the embedding of each word in the transformations under BERT. Underlined words are aspects and corresponding opinion spans.

Conclusion

- We propose three aspect-specific input transformations and methods to leverage these transformations to promote the PLM to pay more attention to the aspect-specific context in two aspect-based sentiment analysis (ABSA) tasks (SC and OE).
- We conduct experiments with standard benchmarks for SC and OE, along with adversarial ones for robustness tests.
- Our models with aspect-specific context modeling achieve the state-of-the-art performance for OE and outperform various strong models for SC.
- The extensive experimental results and further analysis indicated that aspect-specific context modeling can enhance the performance of ABSA.
- Additionally, we craft an *adversarial benchmark for ABSA (advABSA)* to see how aspect-specific modeling can impact model robustness.

- arXiv: https://arxiv.org/pdf/2207.08099.pdf
- Github: https://github.com/BD-MF/ASCM4ABSA

The End

Thanks for your attention.