



Aspect-Specific Context Modeling for Aspect-Based Sentiment Analysis

Fang Ma, Chen Zhang, Bo Zhang, and Dawei Song^(✉)

Beijing Institute of Technology, Beijing, China
{mfang, czhang, bo.zhang, dwsong}@bit.edu.cn

Abstract. Aspect-based sentiment analysis (ABSA) aims at predicting sentiment polarity (SC) or extracting opinion span (OE) expressed towards a given aspect. Previous work in ABSA mostly relies on rather complicated aspect-specific feature induction. Recently, pretrained language models (PLMs), e.g., BERT, have been used as context modeling layers to simplify the feature induction structures and achieve state-of-the-art performance. However, such PLM-based context modeling can be not that aspect-specific. Therefore, a key question is left under-explored: how the aspect-specific context can be better modeled through PLMs? To answer the question, we attempt to enhance aspect-specific context modeling with PLM in a non-intrusive manner. We propose three aspect-specific input transformations, namely aspect companion, aspect prompt, and aspect marker. Informed by these transformations, non-intrusive aspect-specific PLMs can be 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. Extensive experimental results on standard and adversarial benchmarks for SC and OE demonstrate the effectiveness and robustness of the proposed method, yielding new state-of-the-art performance on OE and competitive performance on SC.

Keywords: Aspect-based sentiment analysis · Context modeling · Pretrained language model

1 Introduction

Aspect-based sentiment analysis (ABSA) aims to infer multiple fine-grained sentiments from the same content, with respect to multiple aspects. A fine-grained sentiment in ABSA can be categorized into two forms, i.e., sentiment and opinion. Accordingly, two sub-tasks of ABSA are aspect-based sentiment classification (SC for short) and aspect-based opinion extraction (OE for short). Given an aspect in a sentence, SC aims to predict its sentiment polarity, while OE aims to extract the corresponding opinion span expressed towards the given aspect. Figure 1 shows an example of SC and OE. In the sentence “*The food is tasty but*

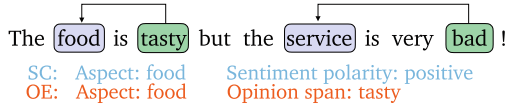


Fig. 1. Example of the SC and OE. The words highlighted in purple represent the given aspects, whereas the words in green represent the corresponding opinion. (Color figure online)

the service is very bad!”, if *food* is the given aspect, SC requires a model to give a **positive** sentiment on *food* while OE requires a model to extract *tasty* as the opinion span for the aspect *food*.

An effective ABSA model typically would require either aspect-specific feature induction or context modeling. Prior work in ABSA largely relies on rather complicated aspect-specific feature induction to achieve a good performance. Recently, pretrained language models (PLMs) have been shown to enhance the state-of-the-art ABSA models due to their extraordinary context modeling ability. However, currently the use of PLMs in these ABSA models is aspect-general, but overlooks two key questions: 1) whether the context modeling of a PLM can be aspect-specific; and 2) whether the aspect-specific context modeling within a PLM can further enhance ABSA.

To address the aforementioned key questions, in this paper, we propose to achieve *aspect-specific context modeling* of PLMs with *aspect-specific input transformations*. In addition to the commonly used aspect-specific input transformation that appends an aspect to a sentence, i.e., **aspect companion**, we propose two more aspect-specific input transformations, namely **aspect prompt** and **aspect marker**, to explicitly mark a concerned aspect in a sentence. Aspect prompt shares a similar idea with aspect companion, except that it appends an aspect-oriented prompt instead of sole aspect description to the sentence. Aspect marker distinguishes itself from the above two by introducing two marker tokens, one before and the other after the aspect. As the proposed input transformations are intended to highlight a specific aspect, they in turn can be leveraged to promote the PLM to pay more attention to the context that is relevant to the aspect. Methodologically, this is achieved with a novel aspect-focused PLM fine-tuning model that is guided by the input transformations and essentially performs a joint context modeling and aspect-specific feature induction.

We conduct extensive experiments on both subtasks of ABSA, i.e., SC and OE, with various standard benchmarking datasets for effectiveness test, along with our crafted adversarial ones for robustness test. Since there are only datasets for robustness tests in SC and is currently no dataset for robustness tests in OE, we propose an adversarial benchmark (advABSA) based on [23]’s datasets and methods. That is, the advABSA benchmark can be decomposed to two parts, where the first part is ARTS-SC for SC reused from [23] and the second part is ARTS-OE for OE crafted by us. The results show that models with aspect-specific context modeling achieve the state-of-the-art performance on OE

and also outperform various strong SC baseline models without aspect-specific modeling. Overall, these results indicate that aspect-specific context modeling for PLMs can further enhance the performance of ABSA.

To better understand the effectiveness of the three input transformations, we carry out a series of further analyses. After injecting aspect-specific input transformations into a sentence, we observe that the model attends to the correct opinion spans. Hence, we expect that a simple model with aspect-specific context modeling yet without needing complicated aspect-specific feature induction would serve as a sufficiently strong approach for ABSA.

2 Related Work

2.1 Aspect-Based Sentiment Classification (SC)

ABSA falls in the broad scope of fine-grained opinion mining. As a sub-task of ABSA, SC determines the sentiment polarity of a given aspect in a sentence and has recently emerged as an active research area with lots of aspect-specific feature induction approaches. These approaches range from memory networks [18, 20], convolutional networks [6, 8, 28], attentional networks [11, 21], to graph-based networks [19, 27]. More recently, PLMs such as BERT [3] and RoBERTa [9], have been applied to SC in a context-encoder scheme [17, 25] and achieved the state-of-the-art performance. However, PLMs in these models are aspect-general. We aim to achieve aspect-specific context modeling with PLMs so that these models can be further improved.

2.2 Aspect-Based Opinion Extraction (OE)

OE is another sub-task of ABSA, first proposed by [4]. It aims to extract from a sentence the corresponding opinion span describing an aspect. Most work in this area treats OE as a sequence tagging task, for which complex methods are developed to capture the interaction between the aspect and the context [4, 5, 22]. More recent models such as TSMSA-BERT [5] and ARGCN-BERT [7], adopt PLMs. In TSMSA-BERT, the multi-head self-attention is utilized to enhance the BERT. ARGCN-BERT uses an attention-based relational graph convolutional network with BERT to exploit syntactic information. We will incorporate our aspect-specific context modeling methods into PLMs to see whether the proposed methods can further improve the OE performance.

3 Aspect-Specific Context Modeling

3.1 Task Description

ABSA (Both SC and OE) requires a pre-given aspect. Formally, a sentence is depicted as $S = \{w_1, w_2, \dots, w_n\}$ that contains n words including the aspect. The aspect $A = \{a_1, a_2, \dots, a_m\}$ is composed of m words. The goal of SC is to find

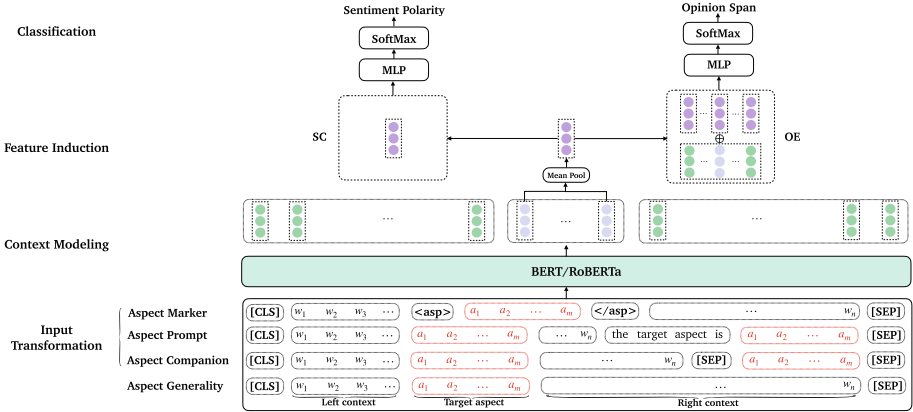


Fig. 2. The architecture of our proposed model based on the three mechanisms.

the sentiment polarity with respect to the given aspect A . OE aims to extract corresponding opinion span based on the given aspect A . Recap the example in Fig. 1 that contains aspect *food*. SC requires a model to give a positive sentiment on *food* and OE requires a model to tag the sentence as $\{0, 0, 0, B, 0, 0, 0, 0, 0, 0, 0, 0, \}$, indicating the opinion span *tasty* for the aspect *food*.

3.2 Overall Framework

Figure 2 shows the structure of our model. Conventionally, an ABSA model consists of four parts: an input layer, a context modeling layer, a feature induction layer, and a classification layer. For aspect-specific context modeling, we first use an aspect-specific transformation to enrich the input. Next, the PLM is applied to get contextualized representations. Then we apply a mean pool operation on the hidden states of the first and last aspect tokens to induct the aspect-specific feature. For SC, we use the aspect-specific feature as the final representation for sentiment classification. For OE, we concatenate the aspect-specific feature and each token’s representation to form the final representation for span tagging.

3.3 Aspect-General Input

The PLM requires a special classification token [CLS] (BERT) or <s> (RoBERTa) be appended to the start of the input sequence, and a separation token [SEP] (BERT) or </s> (RoBERTa) appended to the end of the input sequence. The original input sentence is converted to the format [CLS] + input sequence + [SEP]. We refer to this format as aspect-general input, termed as **aspect generality**. Most previous work uses it for ABSA tasks, and [CLS] is often used for downstream classification, but there is no clear aspect information and no way of knowing which aspect is the focus.

3.4 Aspect-Specific Input Transformations

We propose three aspect-specific input transformations at the input layer to highlight the aspect in the sentence, namely aspect companion, aspect prompt, and aspect marker. We hypothesize that the three transformations can promote the aspect-awareness of PLM and help PLM achieve an effective aspect-specific context modeling.

Aspect Companion. Inspired by BERT’s sentence pair encoding fashion, previous work [24] appends the aspect to the sentence as auxiliary information. Let \hat{S} denote the modified sequence with aspect companion: $\hat{S} = \{[\text{CLS}], w_1, \dots, a_1, \dots, a_m, \dots, w_n, [\text{SEP}], a_1, \dots, a_m, [\text{SEP}]\}$. This formatted sequence can help the PLM effectively model the intra-sentence dependencies between every pair of tokens and further enhance the inter-sentence dependencies between the global context and the aspect.

Aspect Prompt. Inspired by recently popular prompt tuning where some natural language prompts can make the PLM complete a task in a cloze-completion style [1, 15], we here append to the sentence with an aspect-oriented prompt sentence. Let \hat{S} denote the modified sequence with aspect prompt: $\hat{S} = \{[\text{CLS}], w_1, \dots, a_1, \dots, a_m, \dots, w_n, \text{the, target, aspect, is, } a_1, \dots, a_m, [\text{SEP}]\}$. This format sequence prompts the PLM to target at the aimed aspect.

Aspect Marker. Aspect marker inserts markers into the sentence to explicitly mark the boundaries of the concerned aspect. Specifically, we define the markers as two preserved tokens: $\langle \text{asp} \rangle$ and $\langle / \text{asp} \rangle$. We insert them into the input sentence before and after the concerned aspect, to mark the start and end of the given aspect. $\langle \text{asp} \rangle$ indicates the start of the aspect, and $\langle / \text{asp} \rangle$ indicates the end of the aspect. Let \hat{S} denote the modified sequence with aspect marker inserted: $\hat{S} = \{[\text{CLS}], w_1, \dots, \langle \text{asp} \rangle, a_1, \dots, a_m, \langle / \text{asp} \rangle, \dots, w_n, [\text{SEP}]\}$.

The three *aspect-specific input transformations* gain significant improvement in our experiments (Sect. 5), and this strengthens our hypothesis that injecting the aspect marker at the input layer can help the PLM capture aspect-specific contextual information further.

3.5 Context Modeling

Previous PLM-based ABSA work directly adopts the hidden states of the PLM for downstream classification. However, an empirical observation is that the context words close to the aspect are more semantic-relevant to the aspect [12]. In the case, more sentiment information is possibly contained in the aspect’s local context rather than the global context. As a result, the general usage of the hidden states from the PLM loses much local contextual information related to the aspect. With the help of the three input transformations, we obtain the hidden states that incorporate the aspect-oriented local context. Let

$$H = \text{PLM}(\hat{S}) \quad (1)$$

where $H = \{h_1, \dots, h_1^a, \dots, h_m^a, \dots, h_n\}$ represents the sequence of hidden states.

3.6 Feature Induction

As aforementioned, aspect-general feature induction contains the semantic information critical to the whole sentence rather than the given aspect, and the induced aspect-general feature may be aspect-irrelevant when the sentence contains two or more aspects. After getting the global contextual representation H , existing work needs an aspect-specific feature extraction strategy to induce the aspect feature after getting the global contextual representation H . For an enriched aspect-awareness, we adopt the mean pool on the hidden states corresponding to the first and last aspect tokens. Let

$$\hat{H} = \text{MeanPool}([h_1^a, h_m^a]) \quad (2)$$

represent the aspect-specific feature, where h_1^a indicates the hidden state of the first aspect token, and h_m^a indicates the hidden state of the last aspect token. Due to that OE is a token-level classification task, we concatenate the aspect-specific feature \hat{H} and the global contextual representation H as the final aspect-specific contextual representation for tagging.

3.7 Fine-Tuning

After getting the aspect-specific contextual representation \hat{H} , an multi-layered Perceptron (MLP) layer is used to fine-tune the proposed BERT or RoBERTa based model. Then we feed the output to a softmax layer to predict the corresponding label. The training objective is to minimize the cross-entropy loss with \mathcal{L}_2 regularization. Specifically, the optimal parameters θ are obtained from

$$\mathcal{L}(\theta) = - \sum_{i=1}^n \hat{y}_i \log y_i + \lambda \sum_{\theta \in \Theta} \theta^2 \quad (3)$$

where λ is the regularization constant and \hat{y}_i is the predicted label corresponding to ground truth label y_i .

When no input transformation is used, the model is aspect-general and named as PLM-MeanPool and PLM-MeanPool-Concat for SC and OE, respectively. By incorporating the three input transformations, the model becomes more aspect-specific, denoted as +AC (**A**spect **C**ompanion), +AP (**A**spect **P**rompt), and +AM (**A**spect **M**arker) respectively.

4 Experiments

4.1 Datasets

SC Datasets. Following previous work [12], we conduct experiments on two SC benchmarks to evaluate our models’ effectiveness and robustness. One is SemEval 2014 [14] (SEMEVAL), which contains data from laptop (SEM-LAP) and restaurant (SEM-REST) domains; the other is the Aspect Robustness Test Set (ARTS-SC) [23], which is derived from the SEMEVAL dataset. Instances in ARTS-SC are generated with three adversarial strategies. Note that each domain from SEMEVAL consists of separate training and test sets, while each domain from ARTS-SC only contains a test set. Since datasets in SEMEVAL do not contain development sets, 150 instances from the training set in each dataset are randomly selected to form the development set. Table 1 shows the statistics of the SC datasets.

OE Datasets. For datasets used in OE [4,22], the original SEMEVAL benchmark annotates the aspects, but not the corresponding opinion spans, for each sentence. To solve the problem, [4] annotates the corresponding opinion spans for each given aspect in a sentence and removes the cases without explicit opinion spans. We use this variant in our OE experiments.

Since there is currently no robustness test set for OE, we follow [23]’s three adversarial strategies to generate an 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. It is worth noting that this adversarial dataset can also be used for other tasks, e.g., aspect sentiment triplet extraction [13]. Table 2 shows the statistics of the OE datasets. Since these OE datasets do not come with a development set, we randomly split 20% of the training set as validation set.

Table 1. Statistics of SC datasets.

Dataset		#pos.	#neu.	#neg.
Sem-Lap	Train	930	433	800
	Test	341	169	128
	Dev	57	27	66
Sem-Rest	Train	2,094	579	779
	Test	728	196	196
	Dev	70	54	26
Arts-SC-Lap	Test	883	407	587
Arts-SC-Rest	Test	1,953	473	1,104

Table 2. Statistics of OE datasets.

Dataset		#sentences	#aspects
Sem-Lap	Train	1,158	1,634
	Test	343	482
Sem-Rest	Train	1,627	2,643
	Test	500	865
ARTS-OE-Lap	Test	1,002	2,404
ARTS-OE-Rest	Test	2,009	5,743

4.2 Comparative Models and Baselines

We carry out an extensive evaluation of the proposed models (with and without transformation), including *PLM-MeanPool* \pm *AC/AP/AM* for SC, *PLM-MeanPool-Concat* \pm *AC/AP/AM* for OE.

SC Baselines. (a) BERT/RoBERTa-CLS-MLP use the representation of “[CLS]” as a classification feature to fine-tune the BERT/RoBERTa with an MLP layer. (b) AEN-BERT [16] adopts BERT model and attention mechanism to model the relationship between contexts and aspects. (c) LCF-BERT [26] employs Local-Context-Focus design with Semantic-Relative-Distance to discard unrelated sentiment words. (d) BERT/RoBERTa-ASCNN [27] is combined with BERT/RoBERTa and ASCNN model. (e) Roberta-ASGCN [27] use graph convolutional networks to capture the aspect-specific information based on Roberta.

OE Baselines. (a) BERT+Distance-rule [5] is the combination of BERT and Distance-rule. (b) TF-BERT [5] utilizes the average pooling of target word embeddings to represent the target information. (c) SDRN [2] utilizes BERT as the encoder for OE. (d) TSMSA-BERT [5] uses a target-specified sequence labeling method based on multi-head self-attention (TSMSA) to perform OE. (e) ARGCN+BERT [7] adopts the last hidden states of the pretrained BERT as word representations and fine-tune it with the ARGCN model.

4.3 Implementation Details

For fair comparison, we re-produce all baselines based on their open-source codes under the same settings. For experiments with BERT [3] and RoBERTa [9] as the input embeddings, we adopt the BERT-base-uncased model and the RoBERTa-base model as our backbone network, where the learning rate is set to 10^{-5} for SC and $5 * 10^{-5}$ for OE. During all experiments, AdamW [10] is adopted optimizer in our models. The batch size is 64, and the maximal sequence length is 128. It is worth noting that most previous methods did not use the dev set and may have overfitted the test set. We have made a systematic and comprehensive comparison for the first time under the same settings.

4.4 Evaluation Metrics

For standard performance evaluation, each model is trained, validated and tested on the standard datasets. For SC, we use accuracy and macro-averaged F1-score as performance metrics. Following the previous work [4], we adopt F1-score only as the evaluation metric for OE. An opinion extraction is considered correct only when the opinion span predicted is the same as the ground truth. To evaluate a model’s robustness on SC and OE, the model is trained on the standard SEMEVAL datasets and tested on the ARTS-SC and ARTS-OE testsets, respectively. Finally, the experimental results are obtained by averaging five runs with random initialization.

Table 3. Standard and robust experimental results (%) on SC. Our models and better results are bold. 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 (†) compared with our aspect-general base model (BERT-MeanPool/RoBERTa-MeanPool).

Models	SEM-LAP				SEM-REST			
	Standard		Robustness		Standard		Robustness	
	Acc.	F1	Acc.	F1	Acc.	F1	Acc.	F1
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 † 0.22	70.78	67.06 † 0.68	84.71	78.07 † 1.59	78.10	72.38 † 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 † 1.54	80.63	74.03 † 1.30
+AP	81.85	77.91 † 0.23	74.53	70.48	86.43	79.43 † 1.28	80.72	74.09 † 1.36
+AM	82.07 †	78.50 † 0.82	75.90 †	72.59 † 1.38	86.41	79.58 † 1.43	80.88 †	74.04 † 1.31

5 Results and Analysis

5.1 SC Results

Table 3 shows the standard and robustness evaluation results for SC.

Standard Results. Generally, our models with input transformations outperform the baseline models. Before applying the transformations, our base models (BERT/RoBERTa-MeanPool with aspect generality) perform equally good or even better than most baseline models.

Applying the input transformations, especially aspect marker (i.e., +AM), further improves performance significantly. For BERT-based models, the F1-scores of the BERT-MeanPool+AM model are 2.57% and 5.83% higher than AEN-BERT and LCF-BERT respectively on the SEM-REST standard dataset. For RoBERTa-based models, the three transformations are more effective. Specifically, the F1-scores of RoBERTa-MeanPool+AC and RoBERTa-MeanPool+AP improve by up to 1.54% and 1.28% on SEM-REST standard dataset. These results indicate that the proposed input transformations can promote PLMs to achieve effective aspect-specific context modeling.

Among the three transformations, in general AM performs better than AC and AP, indicating that AM is more effective for aspect-specific context modeling in PLMs. While the F1-scores of BERT-MeanPool+AM and

RoBERTa-MeanPool+AM gain improvements by 1.59% and 1.43% on SEM-REST, RoBERTa-MeanPool+AM achieves the terrific results for SC, with F1-score are 78.5% and 79.58% on SEM-LAP and SEM-REST respectively.

Robustness Results. We can see that the performances of the baseline models drop drastically on robustness test sets. In contrast, our models with the transformations are more robust than the baseline models. The most robust model is the RoBERTa-MeanPool+AM, which achieves 72.59% and 74.04% of F1 score on the ARTS-SC-LAP and ARTS-SC-REST robustness test set, representing a 3.21% and 1.48% improvement over the strongest baseline RoBERTa-ASGCN.

The three transformations significantly improve the PLM-MeanPool models' robustness, especially for RoBERTa-MeanPool. Specifically, with AC, AP, and AM, the RoBERTa-MeanPool model's F1-scores are improved by up to 1.30%, 1.36%, and 1.31% on ARTS-SC-REST robustness test set. The model with AM is more robust than the model with AC and AP. These robustness results demonstrate that the transformations can improve our models' robustness.

5.2 OE Results

Table 4 shows the standard and robustness results for OE.

Table 4. Standard and robustness evaluation results (F1-score, %) on OE. The first blocks show the results of the BERT-based baseline models (with *), which are extracted from the published papers [22] and [5]. Note that there were no robustness results of the baseline models in the original published papers, so that we leave them blank. The results of our models are presented in the second and third blocks.

Models	SEM-LAP		SEM-REST	
	Standard	Robustness	Standard	Robustness
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

Standard Results. Before applying the transformations, our base models (PLM-MeanPool-Concat) perform poorly. On the contrary, with the transformations, our models perform significantly better than baselines. Our BERT-based model with the transformations achieves nearly identical results with the current sota model (TSMSA-BERT). With AC, AP, and AM, the F1-scores of the RoBERTa-MeanPool-Concat model are improved by up to 13.04%, 12.89%, and 14.09% on SEM-LAP, respectively. These results demonstrate that the transformations can promote PLMs to achieve effective aspect-specific context modeling for OE. Our RoBERTa-MeanPool-Concat+AM model achieves the new sota result on OE.

Robustness Results. The performances of our base models (PLM-MeanPool-Concat) drop drastically on robustness test set. Their F1-scores are only 39.68% and 38.76% on ARTS-OE-LAP and 44.23% and 56.93% on ARTS-OE-REST. In contrast, with the transformations, our models are more robust, achieving F1 scores up to 73.69% (RoBERTa-MeanPool-Concat+AM) on ARTS-OE-LAP, and 71.61% (RoBERTa-MeanPool-Concat+AP) on ARTS-OE-REST, demonstrating that the transformations can improve our model’s robustness for OE.

Table 5. SC ablation results.

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 6. OE ablation 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

5.3 Ablation Study

To further investigate the effects of the feature induction and the transformations on aspect-specific context modeling of PLMs, we conduct extensive ablation experiments on standard datasets, whose results are included in Table 5 and 6.

Aspect-Specific Feature Induction. For SC and OE, we start with a simple base model that does not use the aspect feature induction component, but using just a context modeling representation after PLM and append an MLP layer (PLM-CLS-MLP for SC, PLM-MLP for OE). After adding back the aspect feature induction, for SC, our PLM-MeanPool models always give a superior

performance than the base model. The F1-scores of BERT-MeanPool are 2.63% higher than BERT-CLS-MLP on SEM-LAP. For OE, our PLM-MeanPool-Concat models perform better than PLM-MLP models. These results demonstrate the effectiveness of the aspect-specific feature induction methods with PLMs.

Aspect-Specific Context Modeling. To investigate the effect of the aspect-specific context modeling with transformations, we add the transformations to the above simple base models. The results show that the transformations bring significant performance improvements, even better than the models with aspect feature induction. Especially the base models with the transformations for OE achieve nearly identical results to BERT/RoBERTa-MeanPool-Concat with transformations. 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.

5.4 Visualization of Attention

To understand the effect of the three transformations, we visualize the attention scores separately offered by our OE model (BERT-MeanPool-Concat) with the transformations, as shown in Fig. 3. The four attention vectors have encoded quite different concerns in the token sequence. We can observe that after applying the transformations, AC, AP, and AM can promote our model to attend to aspect-specific context words and capture the correct opinion spans, thus achieving aspect-specific context modeling in PLM.

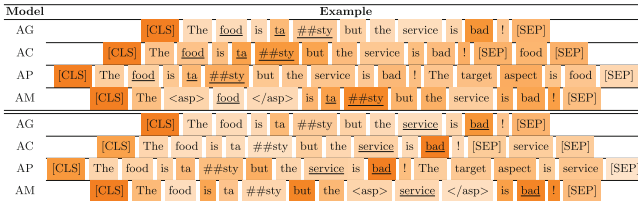


Fig. 3. Attention visualization. Gradient saliency maps for the embedding of each word in the three transformations under BERT architecture. Underlined words are aspects and corresponding opinion spans.

6 Conclusions

In this paper, 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.

Acknowledgements. This research was supported in part by Natural Science Foundation of Beijing (grant number: 4222036) and Huawei Technologies (grant number: TC20201228005).

References

1. Brown, T.B., et al.: Language models are few-shot learners. arXiv preprint [arXiv:2005.14165](https://arxiv.org/abs/2005.14165) (2020)
2. Chen, S., Liu, J., Wang, Y., Zhang, W., Chi, Z.: Synchronous double-channel recurrent network for aspect-opinion pair extraction. In: Proceedings of ACL (2020)
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL (2019)
4. Fan, Z., Wu, Z., Dai, X., Huang, S., Chen, J.: Target-oriented opinion words extraction with target-fused neural sequence labeling. In: Proceedings of NAACL (2019)
5. Feng, Y., Rao, Y., Tang, Y., Wang, N., Liu, H.: Target-specified sequence labeling with multi-head self-attention for target-oriented opinion words extraction. In: Proceedings of NAACL (2021)
6. Huang, B., Carley, K.M.: Parameterized convolutional neural networks for aspect level sentiment classification. In: Proceedings of EMNLP (2018)
7. Jiang, J., Wang, A., Aizawa, A.: Attention-based relational graph convolutional network for target-oriented opinion words extraction. In: Proceedings of EACL (2021)
8. Li, X., Bing, L., Lam, W., Shi, B.: Transformation networks for target-oriented sentiment classification. In: Proceedings of ACL (2018)
9. Liu, Y., et al.: RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint [arXiv:1907.11692](https://arxiv.org/abs/1907.11692) (2019)
10. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: International Conference on Learning Representations (2019)
11. Ma, D., Li, S., Zhang, X., Wang, H.: Interactive attention networks for aspect-level sentiment classification. In: Proceedings of IJCAI (2017)
12. Ma, F., Zhang, C., Song, D.: Exploiting position bias for robust aspect sentiment classification. In: Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, August 2021
13. Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., Si, L.: Knowing what, how and why: a near complete solution for aspect-based sentiment analysis. In: Proceedings of AAAI (2020)
14. Pontiki, M., Papageorgiou, H., Galanis, D., Androutsopoulos, I., Pavlopoulos, J., Manandhar, S.: SemEval-2014 task 4: aspect based sentiment analysis. SemEval 2014 (2014)
15. Schick, T., Schütze, H.: It's not just size that matters: small language models are also few-shot learners. In: Proceedings of NAACL (2021)

16. Song, Y., Wang, J., Jiang, T., Liu, Z., Rao, Y.: Attentional encoder network for targeted sentiment classification. arXiv preprint [arXiv:1902.09314](https://arxiv.org/abs/1902.09314) (2019)
17. Song, Y., Wang, J., Liang, Z., Liu, Z., Jiang, T.: Utilizing BERT intermediate layers for aspect based sentiment analysis and natural language inference. arXiv e-prints (2020)
18. Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. In: Proceedings of EMNLP (2016)
19. Wang, K., Shen, W., Yang, Y., Quan, X., Wang, R.: Relational graph attention network for aspect-based sentiment analysis. In: Proceedings of ACL (2020)
20. Wang, S., Mazumder, S., Liu, B., Zhou, M., Chang, Y.: Target-sensitive memory networks for aspect sentiment classification. In: Proceedings of ACL (2018)
21. Wang, Y., Huang, M., Zhu, X., Zhao, L.: Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of EMNLP (2016)
22. Wu, Z., Zhao, F., Dai, X.Y., Huang, S., Chen, J.: Latent opinions transfer network for target-oriented opinion words extraction. In: Proceedings of AAAI (2020)
23. Xing, X., Jin, Z., Jin, D., Wang, B., Zhang, Q., Huang, X.J.: Tasty burgers, soggy fries: Probing aspect robustness in aspect-based sentiment analysis. In: Proceedings of EMNLP (2020)
24. Xu, H., Liu, B., Shu, L., Philip, S.Y.: Bert post-training for review reading comprehension and aspect-based sentiment analysis. In: Proceedings of NAACL (2019)
25. Yadav, R.K., Jiao, L., Granmo, O.C., Goodwin, M.: Human-level interpretable learning for aspect-based sentiment analysis. In: Proceedings of AAAI (2021)
26. Zeng, B., Yang, H., Xu, R., Zhou, W., Han, X.: LCF: a local context focus mechanism for aspect-based sentiment classification. *Appl. Sci.* **9**, 3389 (2019)
27. Zhang, C., Li, Q., Song, D.: Aspect-based sentiment classification with aspect-specific graph convolutional networks. In: Proceedings of EMNLP (2019)
28. Zhang, C., Li, Q., Song, D.: Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network. In: Proceedings of SIGIR (2019)