Eliminating Contextual Bias in Aspect-based Sentiment Analysis

Ruize An[1], Chen Zhang[1], and Dawei Song[1,2]*

[1] Beijing Institute of Technology, China
{rz.an,czhang,dwsong}@bit.edu.cn
[2] The Open University, United Kingdom

Abstract. Pretrained language models (LMs) have made remarkable achievements in aspect-based sentiment analysis (ABSA). However, it is discovered that these models may struggle in some particular cases (e.g., to detect sentiments expressed towards targeted aspects with only implicit or adversarial expressions). Since it is hard for models to align implicit or adversarial expressions with their corresponding aspects, the sentiments of the targeted aspects would largely be impacted by the expressions towards other aspects in the sentence. We name this phenomenon as contextual bias. To tackle the problem, we propose a flexible aspect-oriented debiasing method (Arde) to eliminate the harmful contextual bias without the need of adjusting the underlying LMs. Intuitively, Arde calibrates the prediction towards the targeted aspect by subtracting the bias towards the context. Favorably, Arde can get theoretical support from counterfactual reasoning theory. Experiments are conducted on SemEval benchmark, and the results show that Arde can empirically improve the accuracy on contextually biased aspect sentiments without degrading the accuracy on unbiased ones. Driven by recent success of large language models (LLMs, e.g., ChatGPT), we further uncover that even LLMs can fail to address certain contextual bias, which yet can be effectively tackled by Arde.

Keywords: aspect-based sentiment analysis · counterfactual inference · implicit sentiment.

1 Introduction

Aspect-based sentiment analysis (ABSA) aims to predict the sentiment expressed towards a particular aspect in a given sentence. Recent advances have preferably employed pretrained language models (LMs) and achieved remarkable gains. Built upon LMs, memory networks [33, 2, 36], convolutional networks [45, 14], attention mechanisms [12, 18], linguistic structures [46, 34, 4], and input transformations [20, 25] have been introduced for aspect-oriented finetuning.

While these aspect-oriented finetuning approaches can largely lift the performance, they may struggle with the so-called contextual bias problem in some

* Corresponding author

Table 1: The accuracy of an existing LM-based ABSA model, namely AspectMarker, on the SemEval Laptop dataset, showing how the model's performance drops in implicit and adversarial cases.

Method	**Normal**	**Implicit**	**Adversarial**
AspectMarker[20]	81.25	71.43	72.25
$\triangle$	**−0.00**	**−9.82**	**−9.00**

particular cases. For example, in a review "*the food here is just great, and the waiter should be more friendly*", the sentiment towards the aspect *waiter* is negative indicated by the implicit expression "*should be more friendly*". However, it can be misjudged as positive due to the explicit context *just great*. As another example, in "*the food here is not bad, but the waiter is awful*", the sentiment towards the aspect *food* is positive but can be misjudged as negative due to the adversarial expressions used. Specifically, although *not just bad* suggests a positive sentiment regarding the *food*, the evident context *awful* in reference to the *waiter* might mislead the overall sentiment assessment towards the *food* as negative. In these cases, the sentiment judgements are largely impacted by the implicit and adversarial expressions towards contextual aspects. We term the afore-discussed phenomena as ***contextual bias***, which can cause remarkable performance degradation in sentiment analysis as demonstrated in Table 1.

Figure 1 provides a quantitative analysis on the impact of contextual bias. Given a review "*My friend had a burger and I had these not wonderful blueberry pancakes*", the sentiment of the aspect *blueberry pancakes* is negative. However, the predicted probability towards it locates more densely at a neutral sentiment. Indeed, the contextual probability for the aspect *blueberry pancakes* leads to a contextual bias that makes its predicted probability move towards a neutral sentiment instead of a negative one.

In this paper, we propose an aspect-oriented debiasing method (Arde) that is aimed to eliminate the harmful contextual bias without any intrusive adjustments to the existing aspect-oriented LM finetuning approaches. Specifically, Arde operates at the LM inference stage and contains three crucial steps. Firstly, for an already finetuned LM for ABSA, Arde obtains the sentiment distribution towards an aspect through LM inference as normal. Then, it gets the sentiment distribution towards the context by LM inference with the aspect-oriented information eliminated. Finally, it induces the calibrated sentiment distribution by conditionally subtracting the bias from the original prediction. Besides, the training-agnostic property of Arde enables universal pluggability to almost all ABSA models.

It is also important to stress that Arde can be viewed as a counterfactual-related instantiation of causal inference [22]. In the language of counterfactuals, the identified contextual bias is a sort of confounding bias and can be counterfactually derived even though it has not ever been seen [27]. As discussed in Section 3.4, the proposed debiasing approach naturally corresponds to spurious

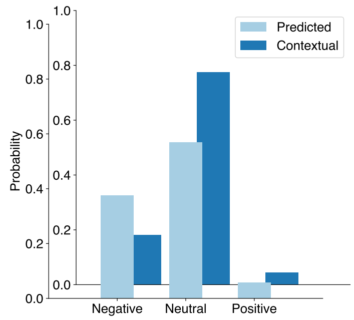


Fig. 1: The predicted probability versus the contextual probability of Aspect-Marker [20] on a case, where the predicted one is biased by the contextual one. The contextual probability is obtained by removing the marker around the target aspect at the input end.

correlation decoupling in the counterfactual theory [22], showing the theoretical soundness of ARDE.

We conduct experiments on the widely used SemEval benchmark [26]. Thanks to previous studies, we can easily distinguish the implicit aspect sentiments [15] and adversarial aspect sentiments [41] from the normal ones. The experimental results demonstrate that ARDE can improve the accuracy on contextually biased aspect sentiments with negligible affect on the unbiased ones.

As a further exploration, we investigate whether the more recently emerged large language models (LLMs, e.g., ChatGPT), which have led to performance breakthroughs in a diverse range of downstream tasks [28], also suffer from the contextual bias problem in ABSA. Our preliminary results indicate that even ChatGPT (at the time when the experiment was carried out) failed to handle the problem, which can be largely alleviated by incorporating ARDE.

Our main contributions can be summarized as follows:

- We discover that existing finetuned LMs can struggle with the contextual bias problem in particular cases, e.g., implicit and adversarial aspect sentiments.
- We design a flexible aspect-oriented debiasing method, ARDE, to eliminate the contextual bias, which is training-agnostic and pluggable to almost all ABSA models.
- We prove that ARDE is theoretically sound from a causal inference perspective and empirically effective on a commonly used benchmark.
- We uncover that LLMs can fail to circumvent contextual bias, but the problem can be effectively tackled by ARDE. To our best knowledge, this is also the very first trial of examining the ability of ChatGPT in the ABSA task.

2 Related Work

2.1 Aspect-based Sentiment Analysis

In recent years, LMs such as BERT [7] and RoBERTa [17], have played a crucial role in NLP. Based on LMs, large performance improvements have been achieved in ABSA [30, 29, 42, 20]. Taking the advantage of LMs, ABSA can be treated as a sequence pair classification task [31, 9] or a reading comprehension task [3, 21]. Moreover, aspect-oriented dependency tree is used to further improve the performance of LMs [47, 46, 40, 5]. Most recently, ABSA is resolved as part of a triplet extraction task to realize a more complete solution [48, 23, 39, 49]. Our work generally falls in this line of approach but specifically concentrates on the contextual bias problem.

2.2 Implicit Aspect Sentiment Analysis

Supervised contrastive pretraining is used to align the representation of implicit sentiment expressions with the corresponding sentiments, and distinguish the implicit slices from the explicit ones in the SemEval benchmark [15]. Furthermore, a knowledge graph is produced to supplement the implicit sentiment expressions and a novel implicit sentiment model is proposed to combine the knowledge enhancements and context features [43]. In addition, structured generation is used for aspect sentiment quadruple extraction to detect implicit sentiment expressions more effectively with a newly-designed quadruple predictor and an encoder-decoder model [24]. Our work utilizes implicit aspect sentiments as one type of contextual bias, and they are also used as instances to measure the performance of our approach in addressing the contextual bias problem.

2.3 Adversarial Aspect Sentiment Analysis

It has been recognized that LMs suffer from significant performance drops on adversarial aspect sentiments [41], and various methods [11, 6, 44, 19] are proposed to improve the robustness of ABSA models. Similarly, a dual-feature extraction module is used to extract aspect-related and aspect-unrelated features, while an aspect-feature distillation module is used to eliminate the interference of aspect-unrelated words [16]. Likewise, our work considers adversarial aspect sentiments as another type of contextual bias and leverages them as a major testbed.

2.4 Debiasing in NLP

Debiasing has been considered as important to improve model robustness in NLP, and a range of methods are proposed for debiasing [35, 27, 10]. Among them, the idea of counterfactual has inspired several debiasing studies [13, 38, 8, 32]. As for ABSA, the bias often refers to the fact that some aspects are more associated with some sentiments rather than others [35]. Such aspect bias can be alleviated by a no-aspect template [1]. In the work [37] that is most related to ours, implicit aspect sentiments are thought to hurt the model robustness and should be intervened by a complex and training-specific instrumental variable model. Differently, our work goes beyond the implicit aspect sentiments, and presents a straightforward and training-agnostic method based on the counterfactual theory to eliminate the contextual bias.

3 Methodology

3.1 Problem Definition

When presented with a sentence $x = (x_1, \dots, a, \dots, x_n)$ (where n denotes the sentence length) and a specific aspect $a = (a_1, \dots, a_m)$ (where m denotes the aspect's length) within the sentence, a fine-tuned Language Model, denoted as $(\mathcal{P}, \mathcal{M})$ for Aspect-Based Sentiment Analysis (ABSA), is essential. This model is tasked with providing a predictive distribution $\bar{y}$ encompassing sentiments (e.g., positive, negative, neutral). Here, $\mathcal{P}$ and $\mathcal{M}$ represent a three-way classifier and a pretrained backbone, respectively. The objective is to align the predicted distribution with the ground truth one-hot distribution y as closely as possible, irrespective of the presence or absence of contextual bias.

3.2 Aspect-oriented Finetuning

Aspect-oriented finetuning of LMs for ABSA can be reduced to two major paradigms: adjusting 1) the input structure of LMs (i.e., $\mathcal{I}$) or 2) the output feature of LMs (i.e., $\mathcal{O}$) for aspect-oriented information. Typical input-based methods include the aspect paired and aspect marked structures [42, 20], while typical output-based methods are based on the aspect averaged and aspect weighted features [5, 47]. These methods are used to make LMs pay mpre attention to the corresponding expressions of the target aspect, and are used in our work for finetuning baselines. An overview of these methods are given in Figure 2a.

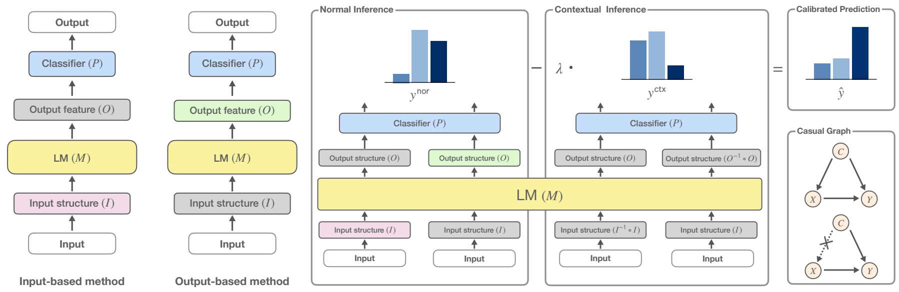
For abstraction, aspect-oriented finetuning can be represented typically as minimizing the cross-entropy loss:

$$\mathcal{L} = -y \log \mathcal{P} \circ \mathcal{O} \circ \mathcal{M} \circ \mathcal{I}(x, a)$$

where $\mathcal{I}$ or $\mathcal{O}$ insert aspect-unaware or aspect-oriented transformations on either the input end or the output end. $\circ$ means sequential function composition.

Aspect Paired Input Structure [42] This input structure is proved to make LMs concentrate more on the targeted aspect by appending the targeted aspect to the sentence:

$$\mathcal{I}(x, a) = [\mathsf{CLS}]\ x_1\ \cdots\ x_n\ [\mathsf{SEP}]\ a_1\ \cdots\ a_m\ [\mathsf{SEP}]$$



(a) (b)

Fig. 2: (a) An overview of aspect-oriented finetuning paradigms. The grey boxes stand for either aspect-unaware input structures or output features. (b) An overview of ARDE and its connections to counterfactual theory.

Aspect Marked Input Structure [20] This input structure is realized by annotating the targeted aspect by placing two markers around the aspect:

$$\mathcal{I}(x,a) = [\mathsf{CLS}]\ x_1\ \cdots\ [\mathsf{M}]\ a_1\ \cdots\ a_m\ [\mathsf{M}]\ \ \cdots\ x_n\ [\mathsf{SEP}]$$

For both input-based methods, the output feature is the last hidden state corresponding to [CLS]:

$$\mathcal{O}(\mathbf{h}, x, a) = \mathbf{h}_{[\mathsf{CLS}]}$$

where $\mathbf{h}$ generally indicates the last hidden states.

Aspect Averaged Output Feature [5] This output feature means taking the average over the last hidden states of the targeted aspect:

$$\mathcal{O}(\mathbf{h}, x, a) = \sum_{k}^{m} 1/m \cdot \mathbf{h}_{a_k}$$

Aspect Weighted Output Feature [47] This output features is derived from the hypothesis that the hidden states locating close to the targeted aspect are more important than the ones locating further away:

$$\mathcal{O}(\mathbf{h}, x, a) = \sum_{k}^{n} d(x_k, a_1, a_m) \cdot \mathbf{h}_{x_k}$$

where d is a function measuring the proximity from any token to the target aspect and it takes the form of:

$$d(x_k, a_1, a_m) = 1 - \min(|x_k, a_1|, |x_k, a_m|)/n$$

where, with abuse of notation, $|p,q|$ denotes the absolute position distance between two tokens.

For both output-based methods, the input structure is the naive sentence:

$$\mathcal{I}(x,a) = [\mathsf{CLS}]\ x_1\ \cdots\ x_n\ [\mathsf{SEP}]$$

3.3 Aspect-oriented Debiasing

Normally the inference with the above finetuned LMs for ABSA will result in errors due to the existence of the contextual bias. As we observe in Figure 1 that the contextual bias can be attributed to the context directly, we propose an aspect-oriented debiasing method which flexibly removes the bias by probability subtraction. Figure 2b provides an overview of our method.

Specifically, the normal inference with LM leads to:

$$y^{\mathsf{nor}} = \mathcal{P} \circ \mathcal{O} \circ \mathcal{M} \circ \mathcal{I}(x, a)$$

In contrast, the contextual inference, where the aspect-oriented information is eliminated and only the context is kept, results in:

$$y^{\mathsf{ctx}} = \mathcal{P} \circ \mathcal{O}^{-1} \circ \mathcal{O} \circ \mathcal{M} \circ \mathcal{I}^{-1} \circ \mathcal{I}(x, a)$$

where $\mathcal{I}^{-1}$ or $\mathcal{O}^{-1}$ is used as inverse function to respectively eliminate the effect of $\mathcal{I}$ or $\mathcal{O}$. Then we can reach a calibrated prediction without bias as:

$$\hat{y} = y^{\mathsf{nor}} - \lambda \cdot y^{\mathsf{ctx}}$$

It is noteworthy that even the subtraction can give negative probability, we can always use it since the LM inference is only concerned with relative magnitudes.

Aspect Paired Input Structure For the Aspect Marked Input Structure, the inverse input structure function is achieved by removing the appended targeted aspect:

$$I^{-1} \circ I(x, a) = [\mathsf{CLS}]\ x_1\ \cdots\ x_n\ [\mathsf{SEP}]$$

Aspect Marked Input Structure Similarly, the inverse input structure function for Aspect Marked Input Structure is implemented by getting rid of the markers around the targeted aspect, which results in an aspect-eliminated input.

Aspect Averaged Output Feature The inverse output feature function for this feature expands the averaging range from the targeted aspect to the whole sentence:

$$\mathcal{O}^{-1} \circ \mathcal{O}(\mathbf{h}, x, a) = \sum_{k}^{n} 1/n \cdot \mathbf{h}_{x_k}$$

Here, the expansion safely erases the aspect-related information.

Aspect Weighted Output Feature The inverse output feature function for this feature drops the proximity weights. Together with the original output feature function, it behaves exactly the same as the one positioned above.

3.4 Connections to Counterfactual Theory

On the theoretical side , we find that ARDE nicely aligns with the counterfactual theory, which essentially tells that the confounding bias can be counter factually conceptualized even though it has never be seen, thus the confounding bias can be described and potentially eliminated [22, 27].

In case of ABSA, the contextual bias is exactly a kind of confounding bias which can be depicted with a causal graph as in Figure 2b (right). In the causal graph, X represents the targeted aspect and its associated sentiment expression, Y represents the ground truth sentiment, and C represents other aspects and their associated sentiment expressions. Two linguistic facts in the causal graph are: 1) X and C are correlated since some aspects would be mentioned concurrently; and 2) C empirically has an impact on Y as sentiment expressions are interconnected and is illustrated in Figure 1. The fork structure $X \leftarrow C \rightarrow Y$ admits C as a confounding bias, which leaves a spurious correlation between X and Y through C.

Fortunately, as the contextual bias can be derived with our tactics, the spurious correlation can be cut off with the subtraction at the probability level.

4 Experiments

4.1 Datasets & Metrics

Our experiments are conducted on the SemEval benchmark. There are generally two domains of data from the benchmark, i.e., one from laptop and the other from restaurant. Each domain consists of both training and test sets. The training part is used to finetune the backbone LMs, and the test part is used to evaluate the model performance. Previous studies have reannotated the test sets to provide implicit test data held out from the original test data [15] and additional adversarial test data [41]. The summary is described in Table 2. We adopt accuracy (Acc.) and F1 scores (F1) as standard measures for evaluation.

Table 2: The summary of data.

Dataset		Positive	Neutral	Negative	Total
Laptop	train	987	460	866	2313
	nor. test	341	169	128	638
	imp. test	36	111	28	175
	adv. test	883	407	587	1877
Restaurant	train	2164	633	805	3602
	nor. test	728	196	196	1120
	imp. test	76	137	54	267
	adv. test	1953	473	1104	3530

4.2 Baselines & Implementation

We compare Arde to the aspect-oriented finetuning baselines, which utilize Bert and Roberta as the backbone LMs. These models are finetuned with learning rate 0.00005 and batch size 64, which are decided by grid searching. The training objective is to minimize the cross-entropy loss with L_2 regularization. We conduct a grid search for the parameter λ over a range from 0 to 1 with a step size of 0.05. This finetuning-searching-predicting process is iterated for five times, and the final results are obtained by averaging the outcomes. For each model, both the mean ($\triangle$) and standard deviation (σ) of the improvements in accuracy and F1 scores are included as part of the results.

Table 3: The results on normal and implicit aspect sentiments (Bert).

Method	Laptop				Restaurant			
	Nor. Acc.	Nor. F1	Imp. Acc.	Imp. F1	Nor. Acc.	Nor. F1	Imp. Acc.	Imp. F1
Aspect Paired Input Structure	78.50	74.04	64.91	58.41	84.32	77.30	64.72	62.81
w/ Arde	78.75	74.44	65.71	58.95	84.37	77.37	64.79	62.87
$\triangle$	**+0.25**	**+0.40**	**+0.80**	**+0.54**	**+0.05**	**+0.07**	**+0.07**	**+0.06**
σ	0.29	0.50	1.06	0.69	0.07	0.09	0.15	0.12
Aspect Marked Input Structure	78.94	75.03	70.63	65.03	84.18	77.44	62.77	61.24
w/ Arde	79.03	75.14	70.74	65.12	84.43	77.96	63.30	61.51
$\triangle$	**+0.09**	**+0.11**	**+0.11**	**+0.09**	**+0.25**	**+0.52**	**+0.53**	**+0.27**
σ	0.19	0.22	0.23	0.18	0.13	0.45	0.30	0.37
Aspect Averaged Output Feature	78.56	74.78	71.43	65.45	85.36	79.36	66.37	63.66
w/ Arde	79.09	75.43	72.57	66.29	85.66	79.97	67.19	64.48
$\triangle$	**+0.53**	**+0.65**	**+1.14**	**+0.84**	**+0.30**	**+0.61**	**+0.82**	**+0.82**
σ	0.26	0.33	0.63	0.57	0.11	0.21	0.49	0.51
Aspect Weighted Output Feature	78.90	75.41	72.00	65.83	84.29	77.80	63.00	60.40
w/ Arde	79.09	75.61	72.11	65.93	84.70	78.46	63.67	61.11
$\triangle$	**+0.10**	**+0.20**	**+0.11**	**+0.10**	**+0.59**	**+0.66**	**+0.67**	**+0.71**
σ	0.15	0.14	0.23	0.18	0.18	0.27	0.44	0.42

4.3 Main Results

Results on Normal Aspect Sentiments As shown in Table 3 and Table 4, we can see that Arde would not affect the performance on normal aspect sentiments and could even bring certain boost. For example, Aspect Weighted Output Feature w/ Arde achieves 0.59 and 0.66 performance gains in terms of Acc. and F1 compared with Aspect Weighted Output Feature on Restaurant in Table 3, and the respective standard deviations of improvement in Acc. and F1 are 0.18 and 0.27. Meanwhile, Aspect Marked Input Structure w/ Arde gives 0.19 and 0.21 absolute performance gains over Aspect Marked Input Structure in terms of Acc. and F1 on Laptop in Table 4, and the standard deviations of improvement are 0.18 and 0.24.

Results on Implicit Aspect Sentiments We find again from Table 3 and Table 4 that Arde can enhance the performance more significantly on implicit aspect

Table 4: The results on normal and implicit aspect sentiments (Roberta).

Method	Laptop				Restaurant			
	Nor. Acc.	Nor. F1	Imp. Acc.	Imp. F1	Nor. Acc.	Nor. F1	Imp. Acc.	Imp. F1
Aspect Paired Input Structure	83.01	80.17	80.45	76.81	87.05	80.53	69.21	68.03
w/ ARDE	83.04	80.21	80.45	76.81	87.12	80.73	69.51	68.41
△	+0.03	+0.04	+0.00	+0.00	+0.07	+0.20	+0.30	+0.38
σ	0.06	0.08	0	0	0.07	0.19	0.28	0.34
Aspect Marked Input Structure	83.86	81.05	83.09	78.95	87.36	81.53	70.56	69.35
w/ ARDE	84.05	81.26	83.43	79.40	87.53	81.79	71.01	69.81
△	+0.19	+0.21	+0.34	+0.45	+0.18	+0.26	+0.45	+0.46
σ	0.18	0.24	0.28	0.37	0.27	0.38	0.60	0.61
Aspect Averaged Output Feature	83.57	80.71	83.43	79.73	86.80	80.55	70.00	69.29
w/ ARDE	83.86	81.13	84.00	80.37	87.02	80.93	71.39	70.15
△	+0.28	+0.41	+0.57	+0.64	+0.21	+0.38	+0.98	+0.86
σ	0.18	0.23	0.63	0.66	0.14	0.29	0.81	0.78
Aspect Weighted Output Feature	83.17	80.09	78.74	73.48	86.84	80.49	70.04	68.91
w/ ARDE	83.20	80.11	78.86	73.57	86.95	80.67	70.26	69.00
△	+0.03	+0.03	+0.11	+0.09	+0.11	+0.18	+0.22	+0.21
σ	0.06	0.05	0.23	0.19	0.07	0.12	0.18	0.17

Table 5: The results on adversarial aspect sentiments (Bert).

Method	Laptop		Restaurant	
	Adv. Acc.	Adv. F1	Adv. Acc.	Adv. F1
Aspect Paired Input Structure	66.02	63.22	77.64	71.62
w/ ARDE	66.59	63.93	77.96	71.98
△	+0.57	+0.71	+0.32	+0.36
σ	0.23	0.36	0.22	0.26
Aspect Marked Input Structure	72.25	69.33	78.47	72.34
w/ ARDE	73.08	70.30	78.82	72.83
△	+0.83	+0.97	+0.35	+0.49
σ	1.22	1.66	0.32	0.48
Aspect Averaged Output Feature	72.32	69.35	77.89	72.48
w/ ARDE	73.35	70.77	78.53	73.32
△	+1.03	+1.42	+0.64	+0.84
σ	1.13	1.53	0.21	0.22
Aspect Weighted Output Feature	70.69	68.03	77.71	72.04
w/ ARDE	71.52	68.90	78.56	73.17
△	+0.83	+0.87	+0.85	+1.13
σ	0.62	0.79	0.62	0.92

sentiments to a large extent. As shown in Table 3, Aspect Averaged Output Feature w/ ARDE improves the Acc. and F1 on Laptop by 1.14 and 0.84 over Aspect Averaged Output Feature, and the standard deviations of improvement are 0.63 and 0.57. The improvement can also be seen on Laptop with a 0.57 and 0.64 boost respectively on Acc. and F1 in Table 4, while the standard deviations of improvement are 0.63 and 0.66. The improvements outweigh those on normal aspect sentiments. The results retrospectively demonstrate that the implicit expressions can lead to contextual bias, which can be alleviated by ARDE.

Results on Adversarial Aspect Sentiments From Table 5 and Table 6, we unearth a similar performance trend on adversarial aspect sentiments. As shown in Table 5, Aspect Weighted Output Feature w/ ARDE improves the Acc. and F1 on Laptop respectively by 0.83 and 0.87 over Aspect Weighted Output Feature. Additionally, the standard deviations of improvement are 0.62 and 0.92. As for

Table 6: The results on adversarial aspect sentiments (Roberta).

Method	**Laptop**		**Restaurant**	
	Adv. Acc.	**Adv. F1**	**Adv. Acc.**	**Adv. F1**
Aspect Paired Input Structure	75.64	73.10	81.69	75.37
w/ ARDE	75.97	73.42	81.83	75.58
$\triangle$	**+0.33**	**+0.32**	**+0.14**	**+0.22**
σ	0.28	0.35	0.10	0.19
Aspect Marked Input Structure	77.44	74.96	81.77	76.12
w/ ARDE	77.75	75.29	82.06	76.49
$\triangle$	**+0.31**	**+0.33**	**+0.29**	**+0.37**
σ	0.16	0.16	0.23	0.33
Aspect Averaged Output Feature	76.51	74.39	81.13	75.01
w/ ARDE	77.11	74.65	81.59	75.71
$\triangle$	**+0.61**	**+0.25**	**+0.46**	**+0.70**
σ	0.29	1.18	0.12	0.17
Aspect Weighted Output Feature	75.23	72.62	80.21	74.20
w/ ARDE	75.65	73.11	80.57	74.76
$\triangle$	**+0.43**	**+0.49**	**+0.36**	**+0.55**
σ	0.25	0.31	0.14	0.25

Aspect Averaged Output Feature w/ ARDE in Table 6, the improvement of the Acc. and F1 on Restaurant are 0.46 and 0.70 over Aspect Averaged Output Feature and the associated standard deviations of improvement are 0.12 and 0.17. The results indicate that Aspect Weighted Output Feature with ARDE exhibits a higher level of stability in handling adversarial aspect sentiments. The results validate that the output-based methods can handle contextual bias more effectively and more stably than the input-based methods.

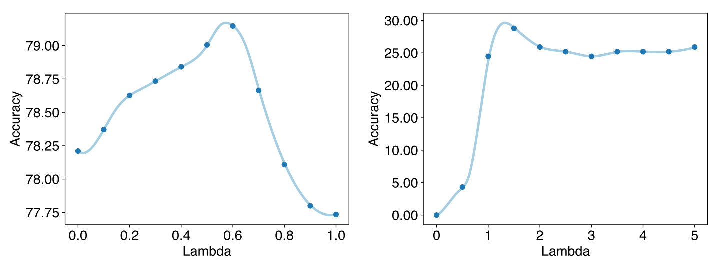


Fig. 3: The impact of λ.

Impact of λ To investigate the impact of λ, we compare the results of Aspect Average Output Feature w/ ARDE on Laptop normal aspect sentiments corresponding to different values of λ. We vary λ from 0 to 1 with a step of 0.05. From Figure 3 (left), we can observe a first-increase-then-decrease phenomenon on Acc., suggesting there is an optimal λ value for ARDE.

Besides, it is not always sufficient for some particular cases to calibrate correctly (i.e., under-calibration) with λ valued from 0 to 1. Yet, λ valued larger than 1 could possibly lead to worse results due to over-calibration on adequately calibrated aspect sentiments. So, we further study the impact of λ ranging from 0 to 5 with a step of 0.5 on probably under-calibrated aspect sentiments, which are wrongly-predicted aspect sentiments (i.e., Acc. is 0 when λ is 0) from the concerned baselines. According to Figure 3 (right), Aspect Average Output Feature w/ ARDE shows an Acc. peak of 37.5 at λ of 1.5, and it also performs well with λ valued from 1.5 to 5.0, maintaining an Acc. of at least 30.0. The results showcase that ARDE with larger λ can have a wider applicable range on more extreme cases, which can't be solved with λ valued from 0 to 1.

4.4 Case Study

Table 7: The attention study. The targeted aspects are underlined. The contextual aspects are underwaved. The groudtruth sentiments are [bracketed]. In these examples, the contextual aspects are paid with more attention, leading to contextual bias.

Example	Contextual Aspect	Targeted Aspect
[CLS] reasonably priced with very stale su ##shi . [SEP]	Pos.	Pos. [Neg.]
[CLS] coffee is a better deal than over ##pr ##ice ##d co ##si sandwiches . [SEP]	Neu.	Neu. [Pos.]
[CLS] go with the specials , and stay away from the salmon . [SEP]	Neg.	Neu. [Pos.]
[CLS] dessert ##s are almost credible : my personal favorite is their tar ##t of the day . [SEP]	Pos.	Pos. [Neg.]
[CLS] it is a not great size and amazing windows 8 included ! [SEP]	Pos.	Pos. [Neg.]
[CLS] :) great product , not great price , great delivery , and great service . [SEP]	Pos.	Pos. [Neg.]
[CLS] it is really easy to use but it is not quick to start up . [SEP]	Pos.	Pos. [Neg.]

Attention Study We select some representative examples to show the attention pattern of contextual bias. Table 7 visualizes the attention of Aspect Average Output Feature without any aspect-oriented information given. We can see that contextual bias arises with varying attentions on different aspects. The unevenness of attention make the aspects that the model pays more attention to play a greater role in the contextual prediction. Therefore, when predicting the sentiments towards the aspects overlooked by the model, they are more likely to be influenced by the aspects of high presence, possibly leading predictions towards an opposite direction.

4.5 Results on ChatGPT

We also apply ARDE to a LLM, namely ChatGPT (v3.5-0315). Different from application of ARDE to aspect-oriented finetuning, we need to give ChatGPT a background about ABSA so that ChatGPT can output the predictive sentiment probability towards the targeted aspect. Then, to evaluate ARDE, we introduce ARDE to calibrate the sentiment probabilities produced from ChatGPT, and

Table 8: ABSA Results with ChatGPT using sampled data on LapTop.

λ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2
Adv.Acc.	79.17	79.17	79.17	79.17	79.17	81.25	82.29	80.21	76.04	66.67	36.46	18.75	14.58

uncover the examples that simply can not be solved by ChatGPT but can be tackled by ARDE.

The evaluation in Table 8 on Laptop adversarial aspect sentiments shows that ARDE can effectively promote the performance on ChatGPT by adjusting the coefficient λ properly. That is, we find that ChatGPT tends to be confused by these particular cases. As shown in Table 9, we can use ARDE to remove the contextual bias, thus yielding correct answers.

Table 9: A Case Study on ChatGPT.

Example		**Predicted Distribution**		
Sentence	**Aspect**	**Pos.**	**Neu.**	**Neg.**
It has all the not expected features and more + plus a wide screen but more than roomy keyboard.	features	Normal		
		0.98	0.01	0.01
		Contextual		
		0.92	0.06	0.02
		Calibrated (λ=1.5)		
		-0.4	-0.08	-0.02

5 Conclusions and Future Work

We have proposed ARDE, a training-agnostic and pluggable method to eliminate the contextual bias in aspect-based sentiment analysis. The approach has been shown theoretically sound from a causal inference perspective, and empirically effective. Another important discovery is that even LLMs can make mistakes when faced with contextual bias, and the problem can be largely alleviated by ARDE. Moving forward, we will broaden the application scope, e.g., to joint aspect extraction and sentiment analysis, and explore the inference costs. Currently, while using ChatGPT, we identified specific instances but were unable to obtain comprehensive results across the entire dataset. In the future, we aim to conduct further testing of ARDE within the state-of-the-art LLM framework using the complete dataset.

Acknowledgements

This work is funded in part by the Natural Science Foundation of China (grant no: 62376027) and Beijing Municipal Natural Science Foundation (grant no: 4222036 and IS23061).

References

1. Cao, J., Liu, R., Peng, H., Jiang, L., Bai, X.: Aspect is not you need: No-aspect differential sentiment framework for aspect-based sentiment analysis. In: Carpuat, M., de Marneffe, M., Ruíz, I.V.M. (eds.) Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022, Seattle, WA, United States, July 10-15, 2022. pp. 1599–1609. Association for Computational Linguistics (2022). https://doi.org/10.18653/v1/2022.naacl-main.115, https://doi.org/10.18653/v1/2022.naacl-main.115
2. Chen, P., Sun, Z., Bing, L., Yang, W.: Recurrent attention network on memory for aspect sentiment analysis. In: Proceedings of the 2017 conference on empirical methods in natural language processing. pp. 452–461 (2017)
3. Chen, S., Wang, Y., Liu, J., Wang, Y.: Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021. pp. 12666–12674. AAAI Press (2021), https://ojs.aaai.org/index.php/AAAI/article/view/17500
4. Chen, X., Sun, C., Wang, J., Li, S., Si, L., Zhang, M., Zhou, G.: Aspect sentiment classification with document-level sentiment preference modeling. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 3667–3677 (2020)
5. Dai, J., Yan, H., Sun, T., Liu, P., Qiu, X.: Does syntax matter? A strong baseline for aspect-based sentiment analysis with roberta. In: Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tür, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., Zhou, Y. (eds.) Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021. pp. 1816–1829. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.naacl-main.146, https://doi.org/10.18653/v1/2021.naacl-main.146
6. Deng, P., Yuan, J., Zhao, Y., Qin, B.: Zero-shot aspect-level sentiment classification via explicit utilization of aspect-to-document sentiment composition. CoRR **abs/2209.02276** (2022). https://doi.org/10.48550/arXiv.2209.02276, https://doi.org/10.48550/arXiv.2209.02276
7. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1423, https://doi.org/10.18653/v1/n19-1423
8. Feng, F., Zhang, J., He, X., Zhang, H., Chua, T.: Empowering language understanding with counterfactual reasoning. In: Zong, C., Xia, F., Li, W., Navigli, R. (eds.) Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021. Findings of ACL, vol. ACL/IJCNLP 2021, pp. 2226–2236. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.findings-acl.196, https://doi.org/10.18653/v1/2021.findings-acl.196

9. Gao, L., Wang, Y., Liu, T., Wang, J., Zhang, L., Liao, J.: Question-driven span labeling model for aspect-opinion pair extraction. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021. pp. 12875–12883. AAAI Press (2021), https://ojs.aaai.org/index.php/AAAI/article/view/17523
10. He, M., Chen, X., Hu, X., Li, C.: Causal intervention for sentiment de-biasing in recommendation. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. pp. 4014–4018 (2022)
11. Hou, X., Qi, P., Wang, G., Ying, R., Huang, J., He, X., Zhou, B.: Graph ensemble learning over multiple dependency trees for aspect-level sentiment classification. In: Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tür, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., Zhou, Y. (eds.) Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021, Online, June 6-11, 2021. pp. 2884–2894. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.naacl-main.229, https://doi.org/10.18653/v1/2021.naacl-main.229
12. Huang, B., Ou, Y., Carley, K.M.: Aspect level sentiment classification with attention-over-attention neural networks. In: Social, Cultural, and Behavioral Modeling: 11th International Conference, SBP-BRiMS 2018, Washington, DC, USA, July 10-13, 2018, Proceedings 11. pp. 197–206. Springer (2018)
13. Kaushik, D., Hovy, E.H., Lipton, Z.C.: Learning the difference that makes A difference with counterfactually-augmented data. In: 8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020. OpenReview.net (2020), https://openreview.net/forum?id=Sklgs0NFvr
14. Li, X., Bing, L., Lam, W., Shi, B.: Transformation networks for target-oriented sentiment classification. In: Gurevych, I., Miyao, Y. (eds.) Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers. pp. 946–956. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/P18-1087, https://aclanthology.org/P18-1087/
15. Li, Z., Zou, Y., Zhang, C., Zhang, Q., Wei, Z.: Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pretraining. In: Moens, M., Huang, X., Specia, L., Yih, S.W. (eds.) Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021, Virtual Event / Punta Cana, Dominican Republic, 7-11 November, 2021. pp. 246–256. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.emnlp-main.22, https://doi.org/10.18653/v1/2021.emnlp-main.22
16. Liu, R., Cao, J., Sun, N., Jiang, L.: Aspect feature distillation and enhancement network for aspect-based sentiment analysis. In: Amigó, E., Castells, P., Gonzalo, J., Carterette, B., Culpepper, J.S., Kazai, G. (eds.) SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, July 11 - 15, 2022. pp. 1577–1587. ACM (2022). https://doi.org/10.1145/3477495.3531938, https://doi.org/10.1145/3477495.3531938

17. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V.: Roberta: A robustly optimized BERT pretraining approach. CoRR **abs/1907.11692** (2019), http://arxiv.org/abs/1907.11692
18. Ma, D., Li, S., Zhang, X., Wang, H.: Interactive attention networks for aspect-level sentiment classification. In: Sierra, C. (ed.) Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017. pp. 4068–4074. ijcai.org (2017). https://doi.org/10.24963/ijcai.2017/568, https://doi.org/10.24963/ijcai.2017/568
19. Ma, F., Zhang, C., Song, D.: Exploiting position bias for robust aspect sentiment classification. In: Zong, C., Xia, F., Li, W., Navigli, R. (eds.) Findings of the Association for Computational Linguistics: ACL/IJCNLP 2021, Online Event, August 1-6, 2021. Findings of ACL, vol. ACL/IJCNLP 2021, pp. 1352–1358. Association for Computational Linguistics (2021). https://doi.org/10.18653/v1/2021.findings-acl.116, https://doi.org/10.18653/v1/2021.findings-acl.116
20. Ma, F., Zhang, C., Zhang, B., Song, D.: Aspect-specific context modeling for aspect-based sentiment analysis. In: Lu, W., Huang, S., Hong, Y., Zhou, X. (eds.) Natural Language Processing and Chinese Computing - 11th CCF International Conference, NLPCC 2022, Guilin, China, September 24-25, 2022, Proceedings, Part I. Lecture Notes in Computer Science, vol. 13551, pp. 513–526. Springer (2022). https://doi.org/10.1007/978-3-031-17120-8_40, https://doi.org/10.1007/978-3-031-17120-8_40
21. Mao, Y., Shen, Y., Yu, C., Cai, L.: A joint training dual-mrc framework for aspect based sentiment analysis. In: Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021. pp. 13543–13551. AAAI Press (2021), https://ojs.aaai.org/index.php/AAAI/article/view/17597
22. Pearl, J.: Causal inference in statistics: An overview (2009)
23. 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: The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020. pp. 8600–8607. AAAI Press (2020), https://ojs.aaai.org/index.php/AAAI/article/view/6383
24. Peper, J., Wang, L.: Generative aspect-based sentiment analysis with contrastive learning and expressive structure. In: Goldberg, Y., Kozareva, Z., Zhang, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2022, Abu Dhabi, United Arab Emirates, December 7-11, 2022. pp. 6089–6095. Association for Computational Linguistics (2022), https://aclanthology.org/2022.findings-emnlp.451
25. Phan, M.H., Ogunbona, P.O.: Modelling context and syntactical features for aspect-based sentiment analysis. In: Proceedings of the 58th annual meeting of the association for computational linguistics. pp. 3211–3220 (2020)
26. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: SemEval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014). pp. 27–35. Association for Computational Linguistics, Dublin, Ireland (Aug 2014). https://doi.org/10.3115/v1/S14-2004, https://aclanthology.org/S14-2004

27. Qian, C., Feng, F., Wen, L., Ma, C., Xie, P.: Counterfactual inference for text classification debiasing. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). pp. 5434–5445 (2021)
28. Qin, C., Zhang, A., Zhang, Z., Chen, J., Yasunaga, M., Yang, D.: Is chatgpt a general-purpose natural language processing task solver? CoRR **abs/2302.06476** (2023). https://doi.org/10.48550/arXiv.2302.06476, https://doi.org/10.48550/arXiv.2302.06476
29. Rietzler, A., Stabinger, S., Opitz, P., Engl, S.: Adapt or get left behind: Domain adaptation through BERT language model finetuning for aspect-target sentiment classification. In: Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. (eds.) Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020. pp. 4933–4941. European Language Resources Association (2020), https://aclanthology.org/2020.lrec-1.607/
30. Song, Y., Wang, J., Jiang, T., Liu, Z., Rao, Y.: Attentional encoder network for targeted sentiment classification. CoRR **abs/1902.09314** (2019), http://arxiv.org/abs/1902.09314
31. Sun, C., Huang, L., Qiu, X.: Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). pp. 380–385. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1035, https://doi.org/10.18653/v1/n19-1035
32. Sun, T., Wang, W., Jing, L., Cui, Y., Song, X., Nie, L.: Counterfactual reasoning for out-of-distribution multimodal sentiment analysis. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 15–23 (2022)
33. Tang, D., Qin, B., Liu, T.: Aspect level sentiment classification with deep memory network. In: Su, J., Carreras, X., Duh, K. (eds.) Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016. pp. 214–224. The Association for Computational Linguistics (2016). https://doi.org/10.18653/v1/d16-1021, https://doi.org/10.18653/v1/d16-1021
34. Tang, H., Ji, D., Li, C., Zhou, Q.: Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In: Proceedings of the 58th annual meeting of the association for computational linguistics. pp. 6578–6588 (2020)
35. Wang, B., Shen, T., Long, G., Zhou, T., Chang, Y.: Eliminating sentiment bias for aspect-level sentiment classification with unsupervised opinion extraction. In: Findings of the Association for Computational Linguistics: EMNLP 2021. pp. 3002–3012 (2021)
36. Wang, S., Mazumder, S., Liu, B., Zhou, M., Chang, Y.: Target-sensitive memory networks for aspect sentiment classification. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (2018)
37. Wang, S., Zhou, J., Sun, C., Ye, J., Gui, T., Zhang, Q., Huang, X.: Causal intervention improves implicit sentiment analysis. In: Calzolari, N., Huang, C., Kim, H., Pustejovsky, J., Wanner, L., Choi, K., Ryu, P., Chen, H., Donatelli, L., Ji,

H., Kurohashi, S., Paggio, P., Xue, N., Kim, S., Hahm, Y., He, Z., Lee, T.K., Santus, E., Bond, F., Na, S. (eds.) Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022. pp. 6966–6977. International Committee on Computational Linguistics (2022), https://aclanthology.org/2022.coling-1.607

38. Wei, T., Feng, F., Chen, J., Wu, Z., Yi, J., He, X.: Model-agnostic counterfactual reasoning for eliminating popularity bias in recommender system. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 1791–1800 (2021)
39. Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., Xia, R.: Grid tagging scheme for aspect-oriented fine-grained opinion extraction. CoRR **abs/2010.04640** (2020), https://arxiv.org/abs/2010.04640
40. Wu, Z., Chen, Y., Kao, B., Liu, Q.: Perturbed masking: Parameter-free probing for analyzing and interpreting BERT. In: Jurafsky, D., Chai, J., Schluter, N., Tetreault, J.R. (eds.) Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020. pp. 4166–4176. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.acl-main.383, https://doi.org/10.18653/v1/2020.acl-main.383
41. Xing, X., Jin, Z., Jin, D., Wang, B., Zhang, Q., Huang, X.: Tasty burgers, soggy fries: Probing aspect robustness in aspect-based sentiment analysis. In: Webber, B., Cohn, T., He, Y., Liu, Y. (eds.) Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020. pp. 3594–3605. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.emnlp-main.292, https://doi.org/10.18653/v1/2020.emnlp-main.292
42. Xu, H., Liu, B., Shu, L., Yu, P.S.: BERT post-training for review reading comprehension and aspect-based sentiment analysis. In: Burstein, J., Doran, C., Solorio, T. (eds.) Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers). pp. 2324–2335. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/n19-1242, https://doi.org/10.18653/v1/n19-1242
43. Xu, M., Wang, D., Feng, S., Yang, Z., Zhang, Y.: KC-ISA: an implicit sentiment analysis model combining knowledge enhancement and context features. In: Calzolari, N., Huang, C., Kim, H., Pustejovsky, J., Wanner, L., Choi, K., Ryu, P., Chen, H., Donatelli, L., Ji, H., Kurohashi, S., Paggio, P., Xue, N., Kim, S., Hahm, Y., He, Z., Lee, T.K., Santus, E., Bond, F., Na, S. (eds.) Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022. pp. 6906–6915. International Committee on Computational Linguistics (2022), https://aclanthology.org/2022.coling-1.601
44. Xu, W., Li, X., Deng, Y., Bing, L., Lam, W.: Peerda: Data augmentation via modeling peer relation for span identification tasks. CoRR **abs/2210.08855** (2022). https://doi.org/10.48550/arXiv.2210.08855, https://doi.org/10.48550/arXiv.2210.08855
45. Xue, W., Li, T.: Aspect based sentiment analysis with gated convolutional networks. In: Gurevych, I., Miyao, Y. (eds.) Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers. pp. 2514–2523. Association for Computational Linguistics (2018). https://doi.org/10.18653/v1/P18-1234, https://aclanthology.org/P18-1234/

46. Zhang, C., Li, Q., Song, D.: Aspect-based sentiment classification with aspect-specific graph convolutional networks. In: Inui, K., Jiang, J., Ng, V., Wan, X. (eds.) Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019. pp. 4567–4577. Association for Computational Linguistics (2019). https://doi.org/10.18653/v1/D19-1464, https://doi.org/10.18653/v1/D19-1464
47. Zhang, C., Li, Q., Song, D.: Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network. In: Piwowarski, B., Chevalier, M., Gaussier, É., Maarek, Y., Nie, J., Scholer, F. (eds.) Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2019, Paris, France, July 21-25, 2019. pp. 1145–1148. ACM (2019). https://doi.org/10.1145/3331184.3331351, https://doi.org/10.1145/3331184.3331351
48. Zhang, C., Li, Q., Song, D., Wang, B.: A multi-task learning framework for opinion triplet extraction. In: Cohn, T., He, Y., Liu, Y. (eds.) Findings of the Association for Computational Linguistics: EMNLP 2020, Online Event, 16-20 November 2020. Findings of ACL, vol. EMNLP 2020, pp. 819–828. Association for Computational Linguistics (2020). https://doi.org/10.18653/v1/2020.findings-emnlp.72, https://doi.org/10.18653/v1/2020.findings-emnlp.72
49. Zhang, C., Ren, L., Ma, F., Wang, J., Wu, W., Song, D.: Structural bias for aspect sentiment triplet extraction. In: Calzolari, N., Huang, C., Kim, H., Pustejovsky, J., Wanner, L., Choi, K., Ryu, P., Chen, H., Donatelli, L., Ji, H., Kurohashi, S., Paggio, P., Xue, N., Kim, S., Hahm, Y., He, Z., Lee, T.K., Santus, E., Bond, F., Na, S. (eds.) Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022. pp. 6736–6745. International Committee on Computational Linguistics (2022), https://aclanthology.org/2022.coling-1.585