Enhancing Implicit Sentiment Learning via the Incorporation of Part-of-Speech for Aspect-based Sentiment Analysis

Junlang Wang, Xia Li*, **Junyi He, Yongqiang Zheng** and **Junteng Ma**
School of Information Science and Technology,
Guangdong University of Foreign Studies, Guangzhou, China
{junlangwang,xiali}@gdufs.edu.cn

Abstract

Implicit sentiment modeling in aspect-based sentiment analysis is a challenging problem due to complex expressions and the lack of opinion words in sentences. Recent efforts focusing on implicit sentiment in ABSA mostly leverage the dependency between aspects and pretrain on extra annotated corpora. We argue that linguistic knowledge can be incorporated into the model to better learn implicit sentiment knowledge. In this paper, we propose a PLM-based, linguistically enhanced framework by incorporating Part-of-Speech (POS) for aspect-based sentiment analysis. Specifically, we design an input template for PLMs that focuses on both aspect-related contextualized features and POS-based linguistic features. By aligning with the representations of the tokens and their POS sequences, the introduced knowledge is expected to guide the model in learning implicit sentiment by capturing sentiment-related information. Moreover, we also design an aspect-specific self-supervised contrastive learning strategy to optimize aspect-based contextualized representation construction and assist PLMs in concentrating on target aspects. Experimental results on public benchmarks show that our model can achieve competitive and state-of-the-art performance without introducing extra annotated corpora.

1 Introduction

Aspect-based Sentiment Analysis (ABSA) aims to identify the sentiment polarities towards specific aspects in sentences. For example, in the sentence "*The dessert is incredible but the service is terrible,*" the sentiment polarities towards the aspects "*dessert*" and "*service*" are *positive* and *negative* respectively.

Previous work on aspect-based sentiment analysis has focused on explicit sentiment expression for specific aspect terms. It means that the sentiment polarities towards the aspects can be explicitly revealed by opinion words. e.g., the sentence "*The dessert is incredible*" contains the opinion word "*incredible*" which carries the positive sentiment towards the corresponding aspect "*dessert*". Many studies have been proposed and achieved promising results towards this task, such as attention mechanism-based methods (Ma et al., 2017; Huang et al., 2018; Zhang et al., 2019a; Wu et al., 2022), graph neural network-based methods (Zhang et al., 2019b; Wang et al., 2020; Liang et al., 2022; Zheng et al., 2023), and pre-trained language model-based methods (Song et al., 2019; Phan and Ogunbona, 2020; Dai et al., 2021; Cao et al., 2022).

However, due to the diversity and flexibility of natural language, sentences containing implicit sentiment expressions are common in human speech. For implicit sentiment, we refer to the recognition of subjective textual units where no polarity markers, opinion words or obvious descriptions are present but people are still able to state whether the text portion under analysis expresses the sentiment (Russo et al., 2015). As shown in Table 1, the four sentences can clearly express the sentiment without any opinion words. Taking the second sentence as an example, no opinion words can be found to determine the sentiment polarities towards the aspects "*food*", but people can still recognize that its polarity is negative. Additionally, we find that some complex expressions, such as factual statements and rhetorical

*Corresponding author: xiali@gdufs.edu.cn

Domain	Example	Polarity
Restaurant	(1) The **waiters** even forget their high-tipping regulars.	negative
	(2) They're a bit more expensive than typical, but then again, so is their **food**.	positive
Laptop	(3) My **voice recording** sounds like interplanetary transmissions in Star Wars.	negative
	(4) Can you buy any laptop that matches the **quality** of a MacBook?	positive

Table 1: Several examples of reviews with implicit sentiment expressions about laptops and restaurants where aspects are marked in bold. The "**Polarity**" column indicates the sentiment polarities of aspects.

techniques, are often used to express implicit sentiment, which always contains complex semantics. For example, sentence (1) and sentence (4) in Table 1 are factual statement and rhetorical question respectively. These complex expressions and the absence of opinion words make it more challenging to detect the implicit sentiment of sentences in the ABSA task.

Few previous studies have paid more attention to the implicit sentiment in ABSA. Among them, Yang et al. (Yang and Li, 2021) propose a local sentiment aggregation paradigm for learning the implicit sentiments in a local sentiment aggregation window. Li et al. (Li et al., 2021b) adopt supervised contrastive pre-training on large-scale sentiment annotated corpora to capture both implicit and explicit sentiment orientation towards aspects in reviews. Their results demonstrate promising performance. However, we argue that the complex implicit expressions can be handled with the help of linguistic knowledge. Motivated by the applications of Parts of Speech (POS) in ABSA (Phan and Ogunbona, 2020; Gong et al., 2020) and opinion mining (Dey and Haque, 2008), we suppose that POS-based linguistic knowledge has the potential to enhance implicit sentiment learning in ABSA. Intuitively, specific POS categories imply the orientation of sentiment polarity. As shown in Figure 1, although the sentence lacks opinion words, the verbs also carry rich sentiment information (Chesley et al., 2006; Nicholls and Song, 2009). The verb "*runs*" states the fact about "*virus scan*" without more related descriptions of this aspect. However, "*flickers*" shows the problem of the aspect "*display screen*". The polarities of the sentiments towards them should be neutral and negative, respectively. Such heuristics motivate us to incorporate POS-based linguistic knowledge into ABSA models for enhancing implicit sentiment prediction.

review:	My	computer	**runs**	a	*virus*	*scan*	but	the	*display*	*screen*	**flickers**
POS:	PRON	NOUN	**VERB**	DET	NOUN	NOUN	CONJ	DET	NOUN	NOUN	**VERB**

Figure 1: Review example with its corresponding POS sequence, marked with Universal POS tags (Petrov et al., 2012). The aspect terms and the verbs are marked in italics and bold.

Inspired by the exploitation of prompts (Li et al., 2021a; Ma et al., 2022) and linguistic knowledge in ABSA (Kiritchenko et al., 2014; Phan and Ogunbona, 2020), we propose a PLM-based, linguistically enhanced framework for aspect-based sentiment analysis that incorporates part-of-speech. We first design a template with POS sequences as PLMs' input. With the multi-head self-attention mechanism, PLMs based on the transformer architecture are able to pay attention to the POS tags and their context information (Vaswani et al., 2017), thereby acquiring potential sentiment knowledge from POS sequences. Considering that the POS sequences are essentially the ordered permutations of the POS tags corresponding to the input sequences and not independent natural language sentences, we leverage token-POS alignment to minimize the semantic impact of POS sequences. In addition, motivated by the applications of contrastive learning to optimize the sentence embeddings derived from BERT (Gao et al., 2021; Yan et al., 2021b; Jiang et al., 2022), an aspect-specific self-supervised contrastive learning strategy is proposed to enhance the construction of contextualized representations, which would focus on aspect-related words in context and the target aspects when handling reviews with multiple aspects. We carry out the experiments on the SemEval 2014 (Pontiki et al., 2014) and Twitter (Dong et al., 2014) benchmark datasets. The experimental results demonstrate the efficacy of our proposed framework.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

The main contributions of this work are as follows:

- We analyze the feasibility of incorporating Part-of-Speech to assist PLMs in modeling implicit sentiment and design an input template for PLMs to focus on both aspect-related contextualized features and POS-based linguistic features.
- We propose the token-POS alignment to reduce the influence of POS sequences on semantics. Additionally, the proposed aspect-specific self-supervised contrastive learning can optimize aspect-based contextualized representations construction and help PLMs concentrate on target aspects.
- Experimental results show the effectiveness of our method, which boosts PLMs to achieve competitive and state-of-the-art performance in ABSA with fewer additional parameters.

2 Related work

In this section, we will briefly review the studies on aspect-based sentiment analysis from three perspectives: methods based on attention mechanisms, graph neural networks (GNNs), and pre-trained language models (PLMs). Then we will introduce implicit sentiment study.

ABSA methods based on attention mechanism. The majority of early attention mechanism-based methods construct the relationship between context and aspects to tackle the ABSA task. Wang et al. (2016) and Ma et al. (2017) equip neural networks with attention mechanisms, promoting the model's ability to identify related information about aspects from input reviews. Li et al. (2018) propose a framework that combines contextual features with word representations. Except for concentrating on the relationship between context and aspects. Zhang et al. (2019b) exploit syntactic features from dependency and mark each word in reviews by proximity values.

ABSA methods based on GNNs. ABSA has demonstrated excellent performance in extracting syntactic features from graph structure since the development of graph neural networks. Sun et al. (2019) and Zhao et al. (2020) use the Graph Convolutional Network (GCN) with the dependency graph to model the dependencies of input sentences. Wang et al. (2020) leverage distances between words in the dependency tree and syntactic tags simultaneously to extract syntactic features by the Graph Attention Network (GAT). Xu et al. (2023) propose to divide sentences into structural scopes according to the results of constituency parsing, which improve the performance of GCN in ABSA.

ABSA methods based on PLMs. The emergence of pre-trained language models in recent years has given ABSA methods a new trend. On the one hand, in order to reduce the gap between pre-training and fine-tuning, numerous works propose sentiment-aware pre-training tasks (Yin et al., 2020; Ke et al., 2020; Fan et al., 2022) based on capturing sentiment semantics and incorporating external knowledge (Baccianella et al., 2010). On the other hand, recent efforts to help PLMs overcome the disadvantages of aspect-aware sentiment perception are flourishing. Cao et al. (2022) remove the sentiment bias of aspect terms and proposes a model trained with differential sentiment loss that is based on the model of Song et al. (2019). Ma et al. (2022) design three aspect-specific input transformations for BERT and RoBERTa that enable the enhancement of aspect-specific context modeling. Moreover, other PLM-based methods solve the ABSA task from the perspective of machine reading comprehension (Xu et al., 2019) and natural language generation (Yan et al., 2021a).

For handling implicit sentiment in ABSA, Li et al. (2021b) propose supervised contrastive pre-training that facilitates BERT in learning sentiment knowledge from large-scale sentiment-annotated corpora. The representation of implicit sentiment expressions is aligned with those of explicit sentiment expressions with the same sentiment polarities through supervised contrastive learning. Yang and Li (2021) build the local sentiment aggregation to model sentiment dependency, which promotes the model's ability to learn implicit sentiment by capturing sentiment information from adjacent aspects. A differentially weighted strategy is also proposed for controlling adjacent aspects that contribute different sentiment information. While these approaches improve the learning and modeling of implicit sentiment in ABSA, external large-scale annotated corpora for encoding adjacent aspects are required. In view of the limitations of these approaches, we propose to leverage POS-based linguistic knowledge to assist PLMs in learning and modeling implicit sentiment in ABSA.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

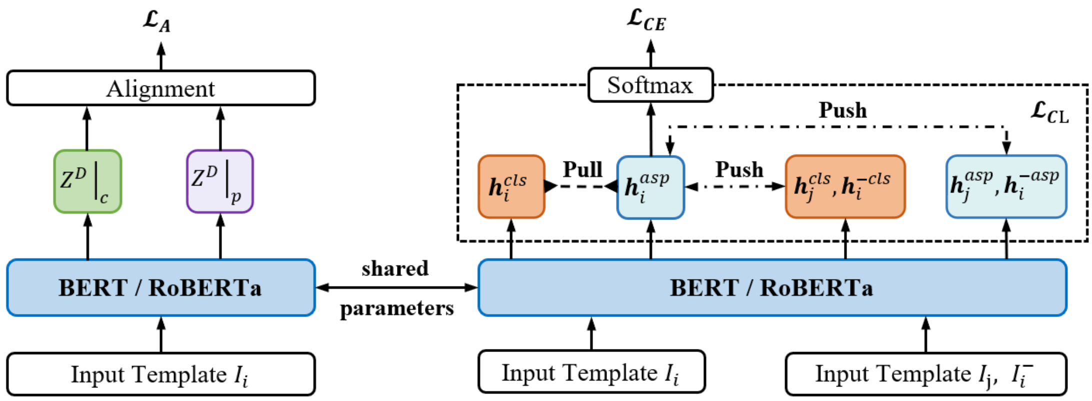

Figure 2: Overall architecture of our proposed framework. In a mini-batch, the input template I_i is derived from the i-th input sentence. I_j, I_i^- represent templates with the other sentence in the mini-batch and the disordered i-th input sentence. $Z^D|_c$ and $Z^D|p$ denote the representations of the input sentences subset and POS sequences subset. $\mathbf{h}_i^{cls}$, $\mathbf{h}_i^{asp}$, $\mathbf{h}_j^{cls}$, $\mathbf{h}_j^{asp}$, $\mathbf{h}_i^{-cls}$ and $\mathbf{h}_i^{-asp}$ are the representations from I_i, I_j and I_i^-, which are elaborated in Section 3.4.

3 Our method

3.1 Overall architecture

As mentioned above, in this paper, we propose a PLM-based linguistically enhanced framework for Aspect-based Sentiment Analysis. Our framework consists of three components: aspect-aware token-POS concatenation, token-POS alignment, and aspect-specific self-supervised contrastive learning. It is expected that POS-based linguistic knowledge will facilitate PLM's learning of implicit sentiment in ABSA. And self-supervised contrastive learning is applied to optimize the representation construction of the target aspect. Our method is shown in Figure 2.

Generally, an input sentence contains one or more aspect terms that correspond to multiple sentiments. In this paper, we focus on the sentiment analysis of a specific aspect. Given a sentence $\mathbf{x} = \{w_1, \ldots, w_t, a_1, \ldots, a_m, w_{t+1}, \ldots, w_n\}$ where w_i indicates the i^{th} word and $asp = \{a_1, \ldots, a_m\}$ denotes the target aspect in $\mathbf{x}$, an input template of our proposed framework I is composed of $\mathbf{x}$ and asp. We will elaborate on the detail of input template I in Section 3.2. The goal of ABSA is to predict the sentiment polarity towards asp according to the sentence $\mathbf{x}$.

3.2 Aspect-aware token-POS concatenation

Motivated by natural language prompt (Brown et al., 2020), we treat the POS sequence as a type of prompt with linguistic knowledge and change the input schema of the PLM. In a mini-batch, for each input sentence $\mathbf{x}_i$, we utilize spaCy[1] to perform part-of-speech tagging on it and combine POS tags into the POS sequence $pos_i = \{p_1, p_2, \ldots, p_n\}$ according to the order of $\mathbf{x}_i$. Instead of concatenating $\mathbf{x}_i$ and pos_i as the input template I_i directly, we additionally append the target aspect term asp_i to I_i following Song et al. (2019), which allows the PLM to capture dependencies between the context and the target aspect:

$$I_i = [CLS] + \mathbf{x}_i + [SEP] + pos_i + [SEP] + asp_i + [SEP] \tag{1}$$

The special tokens "[CLS]" and "[SEP]" of BERT should be "⟨s⟩" and "⟨/s⟩" in RoBERTa. After encoding I_i by BERT or RoBERTa, the pooled representation of asp_i is denoted as $\mathbf{h}_i^{asp} \in \mathbb{R}^{d\times l}$ $(l \geq m)$. Here d is the hidden size of the PLM and l is the length of the tokenized aspect by WordPiece (Wu et al., 2016) or Byte Pair Encoding (Sennrich et al., 2016).

[1]https://spacy.io/

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

3.3 Token-POS alignment

Unlike the discrete templates used in previous research, the POS sequence pos_i is not an independent natural language sentence but the ordered permutation of the POS tags corresponding to the given sentence $\mathbf{x}_i$. To reduce the effects of POS sequences on semantics and promote the interaction of POS sequences and input sentences, we design a token-POS alignment strategy referring to word patch alignment (Kim et al., 2021). As illustrated in Figure 2, in this method, the outputs of the PLM corresponding to the input sentences subset and POS sequences subset in each mini-batch are represented as $Z^D|_c$ and $Z^D|_p$ respectively. After the encoding, $Z^D|_c \in \mathbb{R}^{\mathcal{B}\times k\times d}$ and $Z^D|_p \in \mathbb{R}^{\mathcal{B}\times h\times d}$ can be treated as two different probability distributions, where $\mathcal{B}$ is the mini-batch size, h, k are the lengths of the tokenized input sentence and POS sequence, d is the hidden size. Thus, we convert the alignment into computing the statistical distance between $Z^D|_c$ and $Z^D|_p$, and the alignment score is optimized according to Optimal Transport theory (Peyré et al., 2019). Following such theory, we utilize Wasserstein distance (Vaserstein, 1969) to measure the statistical distance between $Z^D|_c$ and $Z^D|_p$:

$$W_p(Z^D|_c, Z^D|_p) := \mathbb{L}_{M^p}(Z^D|_c, Z^D|_p)^{\frac{1}{p}} \tag{2}$$

where p denotes the p-dimensional Wasserstein distance, $\mathbb{L}_{M^p}$ represents computing Wasserstein distance by Sinkhorn-Knopp algorithm (Knight, 2008) with the constraint of cost matrix $M \in \mathbb{R}^{d\times d}$, and the metric of Sinkhorn-Knopp algorithm is set to the cosine similarity considering the hidden size d of the PLM. Consequently, for $Z^D|_c$ and $Z^D|_p$ within a mini-batch B, the token-POS alignment loss can be defined as:

$$\mathcal{L}_A = \sum_{Z^D|_c, Z^D|_p \in B} W_p(Z^D|_c, Z^D|_p) \tag{3}$$

3.4 Aspect-specific self-supervised contrastive learning

Inspired by the applications of contrastive learning in ABSA (Li et al., 2021b; Liang et al., 2021), we propose to utilize self-supervised contrastive learning to enhance the representation construction of target aspects. According to the aim of contrastive learning (Hadsell et al., 2006), one of the keys is constructing the proper positive instances. Following the previous research in ABSA, both the embedding of the "[CLS]" token (Liang et al., 2021; Zhang et al., 2022) and the aspect features (Dai et al., 2021; Ma et al., 2022) can be used as the final representation for sentiment polarity classification. Hence, those two representations from the same instance can be treated as positives and others from different in-batch instances are taken as negatives. We denote $\mathbf{h}_i^{asp} = f_\theta^{asp}(\mathbf{x}_i)$ where $f_\theta(\cdot)$ represents the encoder. And the embedding of the "[CLS]" token from the same instance is represented as $\mathbf{h}_i^{cls} = f_\theta^{cls}(\mathbf{x}_i)$. Moreover, in order to further leverage the training data and improve the ability of the model to identify the aspect-related context, we construct hard negatives by disordering the input sentence as $\mathbf{x}_i^{dis} = \{w_{t+1}, \ldots, w_n, a_1, \ldots, a_m, w_1, \ldots, w_t\}$. Thus, the input template filled with the disordered input sentence I_i^- is defined as:

$$I_i^- = [CLS] + \mathbf{x}_i^{dis} + [SEP] + pos_i^{dis} + [SEP] + asp_i + [SEP] \tag{4}$$

where pos_i^{dis} is the POS sequence derived from $\mathbf{x}_i^{dis}$. The embedding of the "[CLS]" token and the pooled hidden vector of the aspect term from $\mathbf{x}_i^{dis}$ can be denoted as $\mathbf{h}_i^{-cls} = f_\theta^{cls}(\mathbf{x}_i^{dis})$ and $\mathbf{h}_i^{-asp} = f_\theta^{asp}(\mathbf{x}_i^{dis})$ respectively. Therefore, the aspect-specific self-supervised contrastive loss is defined as ($\mathcal{B}$ is the mini-batch size):

$$\mathcal{L}_{CL} = -\log \frac{e^{sim(\mathbf{h}_i^{asp}, \mathbf{h}_i^{cls})/\tau}}{\sum_{j=1}^{\mathcal{B}} (e^{sim(\mathbf{h}_i^{asp}, \mathbf{h}_j^{cls})/\tau} + e^{sim(\mathbf{h}_i^{asp}, \mathbf{h}_j^{-cls})/\tau} + e^{sim(\mathbf{h}_i^{asp}, \mathbf{h}_j^{-asp})/\tau})} \tag{5}$$

where τ is a temperature hyperparameter and $sim(\mathbf{h}_1, \mathbf{h}_2)$ is the function that computes the cosine similarity between $\mathbf{h}_1$ and $\mathbf{h}_2$.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

3.5 Joint training

Except for applying the two losses mentioned above to optimize the training of our proposed framework, we also use the cross-entropy loss $\mathcal{L}_{CE}$ as the fine-tuning object of the PLM for sentiment polarity prediction:

$$\mathcal{L}_{CE} = -\sum_{i=1}^{\mathcal{B}}\sum_{j=1}^{N} y_i^j \log \hat{y}_i^j + \lambda||\theta||^2 \tag{6}$$

where N is the number of labels, $\mathcal{B}$ is the mini-batch size, λ and θ represent the L_2 regularization and the parameter of the model. As shown in previous studies (Ma et al., 2016), Dropout (Srivastava et al., 2014) may induce inconsistency between the training and inference stages of the model. We argue that such inconsistency will be severe when introducing POS sequences into the input sentences. In order to regularize Dropout, we use the bidirectional Kullback-Leibler (KL) divergence loss $\mathcal{L}_{KL}$ based on R-Drop (Wu et al., 2021) in our models. The overall loss function $\mathcal{L}$ for joint training is:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_1\mathcal{L}_A + \lambda_2\mathcal{L}_{CL} + \alpha\mathcal{L}_{KL} \tag{7}$$

where λ_1 and λ_2 are trainable parameters as the weights of token-POS alignment loss and aspect-specific self-supervised contrastive loss. The coefficient α is a hyperparameter.

4 Experiments

4.1 Datasets

We conduct the experiments using three publicly available benchmark datasets. They are Restaurant and Laptop from SemEval 2014 Task 4 (Pontiki et al., 2014) and Twitter (Dong et al., 2014). The statistics of the three datasets are shown in Table 2. Due to the lack of development sets, 10% of the items from the training sets are randomly selected and treated as development sets. Following previous research, we remove examples with conflicting sentiment polarities.

Dataset	**Positive**		**Neutral**		**Negative**		**Total**	
	Train	Test	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	637	196	807	196	3608	1120
Laptop	994	341	464	169	870	128	2328	638
Twitter	1561	173	3127	346	1560	173	6248	692

Table 2: Statistics on three benchmark datasets of ABSA.

4.2 Implement details

We fine-tune the BERT-base-uncase (Devlin et al., 2019) and RoBERTa-base (Liu et al., 2019) models pre-trained by HuggingFace Transformers (Wolf et al., 2020) and implemented by PyTorch (Paszke et al., 2019). The learning rate is set as 2×10^{-5} and the batch size is 32. We adopt Dropout strategy and the drop probability is adjusted as 0.1. The model is trained with AdamW (Loshchilov and Hutter, 2017) optimizer and the L_2 regularization parameter λ is 10^{-5}. The temperature hyperparameter τ of aspect-specific self-supervised contrastive learning is 0.1. The coefficient α is set as 0.3. Following the work of Chen et al. (2020), we utilize the 2-dimensional Wasserstein distance for token-POS alignment. Since not all of the Universal POS tags exist in the vocabularies of BERT and RoBERTa, we map the tags to their complete names before encoding them to overcome the problem of out-of-vocabulary. We perform our proposed models three runs with different seeds and report their average performance.

4.3 Compared models

In order to demonstrate the effectiveness of our proposed method which can benefit various PLMs in ABSA, we compare the proposed models with several state-of-the-art baselines and models focusing on implicit sentiment in ABSA from the perspectives of BERT-based models and RoBERTa-based models:

Category	Model	Laptop		Restaurant		Twitter	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
BERT	BERT (Devlin et al., 2019)*	77.90	73.37	84.20	76.76	73.70	70.86
	BERT-SPC (Song et al., 2019)	78.99	75.03	84.46	76.98	74.13	72.73
	LCF-BERT (Zeng et al., 2019)*	80.09	76.42	85.65	78.68	74.32	73.32
	BERTAsp (Li et al., 2021b)	78.53	74.07	85.80	78.95	-	-
	BERT+AM (Ma et al., 2022)	76.33	71.93	84.71	78.07	-	-
	IPOS-BERT (Ours)	**80.56**	**76.99**	**85.83**	**79.41**	**76.11**	**74.52**
RoBERTa	RoBERTa (Liu et al., 2019)*	81.97	78.38	87.23	81.00	75.43	74.47
	ASGCN-RoBERTa (Dai et al., 2021)	83.33	80.32	86.87	80.59	76.10	75.07
	RGAT-RoBERTa (Dai et al., 2021)	83.33	79.95	87.52	81.29	75.81	74.91
	LSAP-RoBERTa (Yang and Li, 2021)	83.39	80.47	88.04	82.96	-	-
	RoBERTa+AM (Ma et al., 2022)	82.07	78.50	86.41	79.58	-	-
	IPOS-RoBERTa (Ours)	**83.54**	**80.91**	**88.93**	**83.30**	**77.46**	**76.63**
SCAPT	BERTAsp+SCAPT (Li et al., 2021b)†	82.76	79.15	89.11	83.79	-	-

Table 3: Overall results (%) in three benchmark datasets where the "IPOS-BERT" and "IPOS-RoBERTa" are the proposed models that indicate combining BERT and RoBERTa with our method. The experimental results of the models we reproduced are marked by "∗". For a fair comparison, we mark BERTAsp+SCAPT by "†" and additionally list it in the category "SCAPT" because of its in-domain pre-training and underline its state-of-the-art performance. The best results within other models are highlighted in bold according to different categories.

- **BERT**, **RoBERTa** denote the vanilla BERT and RoBERTa proposed by Devlin et al. (2019) and Liu et al. (2019) respectively. We fine-tune them by ABSA datasets and keep their default settings.
- **BERT-SPC** (Song et al., 2019) transforms the input reviews into sentence-aspect pairs and takes the "[CLS]" token for sentiment polarity classification.
- **LCF-BERT** (Zeng et al., 2019) utilizes the local context focus mechanism to model the relation between global context and local context.
- **BERTAsp** and **BERTAsp+SCAPT** (Li et al., 2021b) are fine-tuned BERT for ABSA. The latter is pre-trained on large-scale annotated corpora by supervised contrastive learning before fine-tuning.
- **ASGCN-RoBERTa**, **RGAT-RoBERTa** are implemented by Dai et al. (2021). They are based on ASGCN (Zhang et al., 2019a) and RGAT (Wang et al., 2020) respectively and RoBERTa is applied including its induced tree and embeddings.
- **LSAP-RoBERTa** (Yang and Li, 2021) aggregates local sentiments by BERT-SPC (Song et al., 2019) and models implicit sentiment by exploiting adjacent aspects' sentiment information.
- **BERT+AM** and **RoBERTa+AM** (Ma et al., 2022) uses the tokens "⟨asp⟩" and "⟨/asp⟩" to mark boundaries of aspects, which promotes PLMs to construct aspect-specific contextualized features.

4.4 Overall results and analysis

The experimental results of the aforementioned compared models and ours are shown in Table 3. Specifically, the accuracy and Macro-F1 score are utilized to evaluate the performance of models. According to the results, we have the following observations:

1) Incorporating linguistic knowledge improves the performance of ABSA models. Compared to the vanilla BERT and RoBERTa, on the one hand, incorporating syntactic knowledge by graph neural networks such as GCN and GAT promotes PLMs to capture the related information about the aspects, which is directly represented as the improvement of ASGCN-RoBERTa and RGAT-RoBERTa. On the other hand, leveraging Part-of-Speech to assist PLMs in modeling implicit sentiment benefits the ABSA task. By incorporating Part-of-Speech and aspect-specific self-supervised contrastive learning, both BERT and RoBERTa improve significantly on three ABSA benchmarks, achieving approximate 2.7%/1.6%/2.4% and 1.6%/1.7%/2.0% performance gains in accuracy as well as 3.6%/2.6%/3.7%

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

Models	Laptop-test		Restaurant-test	
	Accuracy	Accuracy-ISE	Accuracy	Accuracy-ISE
BERT-SPC (Song et al., 2019)	78.99	69.54	84.46	65.54
IPOS-BERT (Ours)	**80.56**	**76.00**	**85.83**	**66.66**
RoBERTa (Liu et al., 2019)	81.97	78.86	87.23	68.54
IPOS-RoBERTa (Ours)	**83.54**	**84.57**	**88.93**	**71.16**

Table 4: Model performance (%) on the Laptop and Restaurant benchmarks and their Implicit Sentiment Expression slices (ISE). The "Accuracy-ISE" column denotes the performance of models on ISE, which is measured by accuracy.

and 2.5%/2.3%/2.2% in Macro-F1 score on Laptop/Restaurant/Twitter benchmarks respectively.

2) Without introducing numerous additional parameters and extra corpora, our model can perform similarly to state-of-the-art models and even outperform them. The proposed model IPOS-RoBERTa has a similar number of parameters (125.2M) to the vanilla RoBERTa-base model (125M). The difference between them lies in the layer for sentiment polarity classification. However, IPOS-RoBERTa can achieve state-of-the-art performance on Laptop and Twitter benchmarks, demonstrating the effectiveness of our method. Unlike LSA_P-RoBERTa and BERTAsp+SCAPT, our method optimizes the fine-tuning of RoBERT to learn implicit sentiment rather than introducing additional parameters for encoding adjacent aspects and extra corpora for pre-training. Specifically, the parameters of the compared models mentioned above are 138.2M and 133.3M respectively[2], indicating millions of parameters are added compared to our proposed model. However, on the test set of the Laptop benchmark, the Macro-F1 score of **IPOS-RoBERTa** is 80.91%, which is 1.76% higher than **BERTAsp+SCAPT** (Macro-F1=79.15%) and 0.44% higher than **LSA_P-RoBERTa** (Macro-F1=80.47%). Though the results of RoBERT-based models on Twitter are not shown in (Yang and Li, 2021), the accuracy and Macro-F1 score of **IPOS-RoBERTa** are 0.55% and 0.73% higher than **LSA_P-DeBERTa** (Accuracy=76.91%, Macro-F1=75.90%), which is based on a progressive PLM called DeBERTa (He et al., 2021). Though the difficulty of improving RoBERTa-based models in ABSA is indicated by Dai et al. (2021), these results prove that POS-based linguistic knowledge and aspect-specific self-supervised contrastive learning are actually beneficial for enhancing the performance of fine-tuned RoBERTa in this task.

4.5 Effectiveness on implicit sentiment learning

Besides conducting extensive experiments on three benchmark datasets mentioned above, we also report the results of the experiment on Implicit Sentiment Expression (ISE) slices of Laptop and Restaurant that are derived from the work of Li et al. (2021b). As shown in Table 4, on both two ISE slices, our proposed models IPOS-BERT and IPOS-RoBERTa outperform compared models based on the same PLMs with them. Though predicting sentiment polarities conveyed by implicit sentiment expressions is challenging, IPOS-RoBERTa's accuracy on ISE slices is higher than that of vanilla RoBERTa by large margins, which indicates the obvious improvement of 5.71% and 2.62%. And the other improvement (6.46% and 1.12%) of Accuracy-ISE can be observed by the comparison of IPOS-BERT and BERT-SPC on the ISE slices of Laptop and Restaurant respectively. Such progresses demonstrates the effectiveness of incorporating POS-based linguistic knowledge for learning implicit sentiment in ABSA.

5 Discussion

5.1 Ablation study

Considering that each component of the proposed framework plays a different role as well as the temperatures contribute variously, extensive ablation experiments are conducted on the Laptop benchmark and results are shown in Table 5. We find that removing the token-POS alignment degrades the proposed

[2]The statistics of parameters are derived from open-source repositories released by Yang and Li (2021) and Li et al. (2021b).

Model Variant	Laptop	
	Accuracy	Macro-F1
IPOS-RoBERTa	83.54	80.91
w/o $\mathcal{L}_A$	81.97	78.80
w/o $\mathcal{L}_{CL}$	82.29	78.85
w/o $\mathcal{L}_{KL}$	82.60	79.25

(a) The ablation study of different components

Model Variant	Laptop	
	Accuracy	Macro-F1
IPOS-RoBERTa	83.54	80.91
$\tau = 0.01$	81.03	77.77
$\tau = 0.05$	83.23	80.65
$\tau = 0.5$	83.07	79.37

(b) The ablation study of different temperatures

Table 5: Ablation studies of different components and temperatures on the Laptop benchmark (%). "w/o $\mathcal{L}_A$, $\mathcal{L}_{CL}$, $\mathcal{L}_{KL}$" indicates the models without token-POS alignment, aspect-specific self-supervised contrastive learning and R-Drop respectively. In the ablation study of temperatures (τ), we compare the original setting ($\tau = 0.1$) with three variants.

Example and POS Sequence	RoBERTa	BERTAsp*	Ours
However, I can refute that OSX is FAST. ADV PRON AUX VERB SCONJ PROPN AUX ADJ	Pos (×)	Pos (×)	Neg (✓)
Fan only comes on when you are playing a game. NOUN ADV VERB ADP SCONJ PRON AUX VERB DET NOUN	Neg, Neu (×), (✓)	Neu, Neu (✓), (✓)	Neu, Neu (✓), (✓)
It has so much more speed and the screen is very sharp. PRON VERB ADV ADV ADJ NOUN CCONJ DET NOUN AUX ADV ADJ	Pos, Pos (✓), (✓)	Pos, Neg (✓), (×)	Pos, Pos (✓), (✓)
I did swap out the hard drive for a Samsung 830 SSD which I highly recommend. PRON AUX VERB ADP DET NOUN ADP DET PROPN PRON PRON ADV VERB	Neu, Neu (✓), (×)	Neu, Neu (✓), (×)	Neu, Pos (✓), (✓)

Table 6: A case study in the domain of laptops. For each case example, the original review and its POS sequence are shown. The model marked by "$*$" denotes BERTAsp+SCAPT proposed by Li et al. (2021b) and the aspect terms are underlined. We use "Pos, Neu, Neg" to indicate three sentiment polarities ("Positive, Neutral, Negative"). The correct predictions are associated with the symbol "✓" and the wrong predictions are marked with "×".

model drastically and even leads to the suboptimal performance of the proposed model, which is similar to that of the vanilla RoBERTa. We suppose that the POS sequences imported from the external parser affect contextual semantics without the token-POS alignment (Similar visual examples are shown in the rows of "RoBERTa (with POS)" in Figure 3). Thus, though keeping the aspect-specific self-supervised contrastive learning and R-Drop, their effects are obscure while importing POS sequences directly. Such degradation indicates the importance of incorporating Part-of-Speech knowledge properly. Another noticeable performance degradation is caused by the absence of aspect-specific self-supervised contrastive learning since it promotes the model to concentrate on the target aspects. Similarly, our model benefits from R-Drop (Wu et al., 2021) due to the regularization of the predictions.

Moreover, in order to investigate the influence of different temperatures, we set the temperature $\tau \in \{0.01, 0.05, 0.1, 0.5\}$ and keep other settings of our model. Compared to the carefully tuned temperature ($\tau = 0.1$), the other lead to different degrees of impact. It is worth noting that an extremely small temperature ($\tau = 0.01$) causes an obvious drop in the performance, which makes the model focus much on hard negatives (Wang and Liu, 2021). However, a high temperature is also inappropriate. Specifically, both the accuracy and the Marco F1 score of the proposed model trained with a high temperature ($\tau = 0.5$) are lower than those of the model with a carefully tuned temperature by large margins.

5.2 Case study

To verify the effectiveness of our method, we select several cases in the laptop domain that contain implicit sentiment expressions, as shown in Table 6. According to these cases, the capabilities of modeling implicit sentiment and capturing syntactic features are demanded. Hence, BERTAsp+SCAPT (Li et al.,

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

Model	Case Visualization	Asp
IPoS-RoBERTa	<s> However I can refute that OSX is FAST </s> ADV PRON AUX VERB SCONJ PROPN AUX ADJ </s> OSX </s>	OSX
RoBERTa (with PoS)	<s> However I can refute that OSX is FAST </s> ADV PRON AUX VERB SCONJ PROPN AUX ADJ </s> OSX </s>	OSX
RoBERTa	<s> However I can refute that OSX is FAST </s> OSX </s>	OSX
IPoS-RoBERTa	<s> I will not be using that slot again </s> PRON AUX PART AUX VERB DET NOUN ADV </s> slot </s>	slot
RoBERTa (with PoS)	<s> I will not be using that slot again </s> PRON AUX PART AUX VERB DET NOUN ADV </s> slot </s>	slot
RoBERTa	<s> I will not be using that slot again </s> slot </s>	slot

Figure 3: Visualization of two selected cases. Both two target aspects are expressed by implicit sentiment. The gradient saliency maps (Simonyan et al., 2014) for the embedding of input tokens are shown, including the words and corresponding POS tags. For each token, the darker color denotes the higher gradient saliency score. The "**Asp**" column indicates the aspect terms.

2021b) and RoBERTa (Liu et al., 2019) are chosen as strong compared models for the case study. Following the comparison results, both BERTAsp+SCAPT and RoBERTa fail to correctly predict all the case examples. For example, RoBERTa wrongly comprehends the semantics of the second review and predicts the sentiment polarity towards "**fan**" as negative, which is represented by implicit sentiment expression. Additionally, for the aspect "**screen**" in the third case, BERTAsp+SCAPT mistakes the opinion "sharp" and incorrectly infers the corresponding polarity as negative. However, when given some complicated cases carrying multiple aspects and intricate implicit sentiment, both of them improperly capture the aspect-related contextualized features such as the aspects "**OSX**" and "**Samsung 830 SSD**" in the first and the last cases.

Owing to the POS-based linguistic knowledge, the proposed IPOS-RoBERTa model can precisely predict all aforementioned cases. We suppose that POS sequences encourage the model to learn implicit sentiment and distinguish sentiment expressions about different aspects, as suggested by the good performance of our proposed model. For the first case, the adjective "FAST" is related to the aspect "**OSX**" from the view of syntax but it implies the contrary sentiment polarity due to the verb "refute", which helps to perceive the implicit sentiment. Moreover, when inferring multiple aspects "**hard drive**" and "**Samsung 830 SSD**" in the same sentence, IPOS-RoBERTa can distinguish the related information about them and predict the correct sentiment polarity towards "**Samsung 830 SSD**".

5.3 Visualization

Since it seems that the effect of appending POS tags to the input tokens is intricate, we visualize the gradient saliency scores of the embedding of input templates for two selected cases, which can be employed for model interpretation (Li et al., 2016). As shown in Figure 3, we compare our model with two backbones and keep the setting that appends the aspect terms to the input sequences for all of them. However, "RoBERTa (with POS)" denotes only employing aspect-aware token-POS concatenation to RoBERTa but ignoring the token-POS alignment and "RoBERTa" indicates the vanilla RoBERTa model.

In the first case, the words "refute" and "FAST" are assigned different saliency scores among the three models, signifying these words differently contributing to the predictions. Compared to another two models, we suppose that our model pays more attention to such important words in comprehending the semantics. Furthermore, due to the token-POS alignment, our model can distinguish the importance of

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

different POS tags instead of treating them equally. Similarly, though the three selected models focus on the word "not", the neglect of the verb "using" leads to incorrect predictions of sentiment polarity towards the aspect "slot". In contrast, our model can precisely capture essential words and their POS for prediction, demonstrating the effect of aspect-aware token-POS concatenation and token-POS alignment.

6 Conclusion

In this paper, we propose a PLM-based linguistically enhanced framework for aspect-based sentiment analysis based on the analysis of the feasibility of incorporating Part-of-Speech into the ABSA task. Using POS-based linguistic knowledge, our method optimizes the PLMs' fine-tuning for implicit sentiment capturing. Aspect-specific self-supervised contrastive learning allows the model to concentrate on target aspects when handling sentences containing multiple aspect terms. Extensive experiments show that our proposed model can achieve competitive and state-of-the-art performance relative to baseline models without introducing extra corpora. Although the introduction of POS as linguistic knowledge can effectively improve the enhancement of implicit sentiment detection in ABSA, there are still limitations. If there are difficulties in deriving precise POS sequences in low-resource settings, the POS-based solution might not provide sufficient information. Further research can investigate approaches for integrating various linguistic knowledge into models for learning implicit sentiment without external sources.

Acknowledgements

This work is supported by National Natural Science Foundation of China (No. 61976062).

References

Stefano Baccianella, Andrea Esuli, Fabrizio Sebastiani, et al. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, volume 10, pages 2200–2204.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Jiahao Cao, Rui Liu, Huailiang Peng, Lei Jiang, and Xu Bai. 2022. Aspect is not you need: No-aspect differential sentiment framework for aspect-based sentiment analysis. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1599–1609, Seattle, United States, July. Association for Computational Linguistics.

Yen-Chun Chen, Linjie Li, Licheng Yu, Ahmed El Kholy, Faisal Ahmed, Zhe Gan, Yu Cheng, and Jingjing Liu. 2020. Uniter: Universal image-text representation learning. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXX*, pages 104–120. Springer.

Paula Chesley, Bruce Vincent, Li Xu, and Rohini K Srihari. 2006. Using verbs and adjectives to automatically classify blog sentiment. *Training*, 580(263):233.

Junqi Dai, Hang Yan, Tianxiang Sun, Pengfei Liu, and Xipeng Qiu. 2021. Does syntax matter? a strong baseline for aspect-based sentiment analysis with RoBERTa. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1816–1829, Online, June. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Lipika Dey and S K Mirajul Haque. 2008. Opinion mining from noisy text data. In *Proceedings of the Second Workshop on Analytics for Noisy Unstructured Text Data*, AND '08, page 83–90, New York, NY, USA. Association for Computing Machinery.

Li Dong, Furu Wei, Chuanqi Tan, Duyu Tang, Ming Zhou, and Ke Xu. 2014. Adaptive recursive neural network for target-dependent Twitter sentiment classification. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 49–54, Baltimore, Maryland, June. Association for Computational Linguistics.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

Shuai Fan, Chen Lin, Haonan Li, Zhenghao Lin, Jinsong Su, Hang Zhang, Yeyun Gong, JIan Guo, and Nan Duan. 2022. Sentiment-aware word and sentence level pre-training for sentiment analysis. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 4984–4994, Abu Dhabi, United Arab Emirates, December. Association for Computational Linguistics.

Tianyu Gao, Xingcheng Yao, and Danqi Chen. 2021. SimCSE: Simple contrastive learning of sentence embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6894–6910, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Chenggong Gong, Jianfei Yu, and Rui Xia. 2020. Unified feature and instance based domain adaptation for aspect-based sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7035–7045.

Raia Hadsell, Sumit Chopra, and Yann LeCun. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1735–1742. IEEE.

Pengcheng He, Jianfeng Gao, and Weizhu Chen. 2021. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. *arXiv preprint arXiv:2111.09543*.

Binxuan Huang, Yanglan Ou, and Kathleen M Carley. 2018. 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*, pages 197–206. Springer.

Ting Jiang, Jian Jiao, Shaohan Huang, Zihan Zhang, Deqing Wang, Fuzhen Zhuang, Furu Wei, Haizhen Huang, Denvy Deng, and Qi Zhang. 2022. PromptBERT: Improving BERT sentence embeddings with prompts. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 8826–8837, Abu Dhabi, United Arab Emirates, December. Association for Computational Linguistics.

Pei Ke, Haozhe Ji, Siyang Liu, Xiaoyan Zhu, and Minlie Huang. 2020. Sentilare: Sentiment-aware language representation learning with linguistic knowledge. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6975–6988.

Wonjae Kim, Bokyung Son, and Ildoo Kim. 2021. Vilt: Vision-and-language transformer without convolution or region supervision. In *International Conference on Machine Learning*, pages 5583–5594. PMLR.

Svetlana Kiritchenko, Xiaodan Zhu, Colin Cherry, and Saif Mohammad. 2014. Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014)*, pages 437–442.

Philip A Knight. 2008. The sinkhorn–knopp algorithm: convergence and applications. *SIAM Journal on Matrix Analysis and Applications*, 30(1):261–275.

Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky. 2016. Visualizing and understanding neural models in NLP. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 681–691, San Diego, California, June. Association for Computational Linguistics.

Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. *arXiv preprint arXiv:1805.01086*.

Chengxi Li, Feiyu Gao, Jiajun Bu, Lu Xu, Xiang Chen, Yu Gu, Zirui Shao, Qi Zheng, Ningyu Zhang, Yongpan Wang, et al. 2021a. Sentiprompt: Sentiment knowledge enhanced prompt-tuning for aspect-based sentiment analysis. *arXiv preprint arXiv:2109.08306*.

Zhengyan Li, Yicheng Zou, Chong Zhang, Qi Zhang, and Zhongyu Wei. 2021b. Learning implicit sentiment in aspect-based sentiment analysis with supervised contrastive pre-training. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 246–256, Online and Punta Cana, Dominican Republic, November. Association for Computational Linguistics.

Bin Liang, Wangda Luo, Xiang Li, Lin Gui, Min Yang, Xiaoqi Yu, and Ruifeng Xu. 2021. Enhancing aspect-based sentiment analysis with supervised contrastive learning. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 3242–3247.

Bin Liang, Hang Su, Lin Gui, Erik Cambria, and Ruifeng Xu. 2022. Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowledge-Based Systems*, 235:107643.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Xuezhe Ma, Yingkai Gao, Zhiting Hu, Yaoliang Yu, Yuntian Deng, and Eduard Hovy. 2016. Dropout with expectation-linear regularization. *arXiv preprint arXiv:1609.08017*.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. *arXiv preprint arXiv:1709.00893*.

Fang Ma, Chen Zhang, Bo Zhang, and Dawei Song. 2022. Aspect-specific context modeling for aspect-based sentiment analysis. In *Natural Language Processing and Chinese Computing: 11th CCF International Conference, NLPCC 2022, Guilin, China, September 24–25, 2022, Proceedings, Part I*, pages 513–526.

Chris Nicholls and Fei Song. 2009. Improving sentiment analysis with part-of-speech weighting. In *2009 International Conference on Machine Learning and Cybernetics*, volume 3, pages 1592–1597. IEEE.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Slav Petrov, Dipanjan Das, and Ryan McDonald. 2012. A universal part-of-speech tagset. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2089–2096, Istanbul, Turkey, May. European Language Resources Association (ELRA).

Gabriel Peyré, Marco Cuturi, et al. 2019. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607.

Minh Hieu Phan and Philip O. Ogunbona. 2020. Modelling context and syntactical features for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3211–3220, Online, July. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland, August. Association for Computational Linguistics.

Irene Russo, Tommaso Caselli, and Carlo Strapparava. 2015. Semeval-2015 task 9: Clipeval implicit polarity of events. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pages 443–450.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725.

K Simonyan, A Vedaldi, and A Zisserman. 2014. Deep inside convolutional networks: visualising image classification models and saliency maps. In *Proceedings of the International Conference on Learning Representations (ICLR)*. ICLR.

Youwei Song, Jiahai Wang, Tao Jiang, Zhiyue Liu, and Yanghui Rao. 2019. Attentional encoder network for targeted sentiment classification. *arXiv preprint arXiv:1902.09314*.

Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958.

Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 5679–5688.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

Leonid Nisonovich Vaserstein. 1969. Markov processes over denumerable products of spaces, describing large systems of automata. *Problemy Peredachi Informatsii*, 5(3):64–72.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Feng Wang and Huaping Liu. 2021. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2495–2504.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615.

Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. Relational graph attention network for aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October. Association for Computational Linguistics.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Lijun Wu, Juntao Li, Yue Wang, Qi Meng, Tao Qin, Wei Chen, Min Zhang, Tie-Yan Liu, et al. 2021. R-drop: Regularized dropout for neural networks. *Advances in Neural Information Processing Systems*, 34:10890–10905.

Haiyan Wu, Zhiqiang Zhang, Shaoyun Shi, Qingfeng Wu, and Haiyu Song. 2022. Phrase dependency relational graph attention network for aspect-based sentiment analysis. *Knowledge-Based Systems*, 236:107736.

Hu Xu, Bing Liu, Lei Shu, and Philip Yu. 2019. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1.

Lvxiaowei Xu, Xiaoxuan Pang, Jianwang Wu, Ming Cai, and Jiawei Peng. 2023. Learn from structural scope: Improving aspect-level sentiment analysis with hybrid graph convolutional networks. *Neurocomputing*, 518:373–383.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021a. A unified generative framework for aspect-based sentiment analysis. 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)*, pages 2416–2429.

Yuanmeng Yan, Rumei Li, Sirui Wang, Fuzheng Zhang, Wei Wu, and Weiran Xu. 2021b. Consert: A contrastive framework for self-supervised sentence representation transfer. 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)*, pages 5065–5075.

Heng Yang and Ke Li. 2021. Improving implicit sentiment learning via local sentiment aggregation. *arXiv e-prints*, pages arXiv–2110.

Da Yin, Tao Meng, and Kai-Wei Chang. 2020. Sentibert: A transferable transformer-based architecture for compositional sentiment semantics. *arXiv preprint arXiv:2005.04114*.

Biqing Zeng, Heng Yang, Ruyang Xu, Wu Zhou, and Xuli Han. 2019. Lcf: A local context focus mechanism for aspect-based sentiment classification. *Applied Sciences*, 9(16):3389.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019a. Aspect-based sentiment classification with aspect-specific graph convolutional networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4568–4578.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.

Chen Zhang, Qiuchi Li, and Dawei Song. 2019b. Syntax-aware aspect-level sentiment classification with proximity-weighted convolution network. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 1145–1148.

Kai Zhang, Kun Zhang, Mengdi Zhang, Hongke Zhao, Qi Liu, Wei Wu, and Enhong Chen. 2022. Incorporating dynamic semantics into pre-trained language model for aspect-based sentiment analysis. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3599–3610, Dublin, Ireland, May. Association for Computational Linguistics.

Pinlong Zhao, Linlin Hou, and Ou Wu. 2020. Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification. *Knowledge-Based Systems*, 193:105443.

Yongqiang Zheng, Xia Li, and Jian-Yun Nie. 2023. Store, share and transfer: Learning and updating sentiment knowledge for aspect-based sentiment analysis. *Information Sciences*, 635:151–168.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 786-800, Harbin, China, August 3 - 5, 2023.