





# Prompt-Oriented Fine-Tuning Dual Bert for Aspect-Based Sentiment Analysis

Wen Yin<sup>1,2</sup> , Yi Xu<sup>1,2</sup> , Cencen Liu<sup>1,2</sup>, Dezhang Zheng<sup>1,2</sup>, Qi Wang<sup>3</sup>,  
and Chuanjie Liu<sup>3</sup>

<sup>1</sup> School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

yinwenok@std.uestc.edu.cn xuyi0421@uestc.edu.cn

<sup>2</sup> Trusted Cloud Computing and Big Data Key Laboratory of Sichuan Province, Chengdu 611731, China

<sup>3</sup> Chengdu Jiuzhou Electronic Information System Co., Ltd., Chengdu, China

**Abstract.** Aspect-Based Sentiment Analysis (ABSA) is a fine-grained sentiment analysis task that aims to predict sentiment polarity towards a specific aspect occurring in the given sentence. Recently, pre-trained language models such as BERT have shown great progress in this regard. However, due to the mismatch between pre-training and fine-tuning, dealing with informal expressions and complex sentences is facing challenges and it is worthwhile devoting much effort to this. To tackle this, in this paper, we propose a **Prompt-oriented Fine-tuning Dual BERT (PFDualBERT)** model that considers the complex semantic relevance and the scarce data samples simultaneously. To reduce the impact of such mismatches, we design a ProBERT influenced by the idea of prompt Learning. Specifically, we design a SemBERT module to capture semantic correlations. We refit SemBERT with aspect-based self-attention. The experimental results on three datasets certify that our PFDualBERT model outperforms state-of-the-art methods, and our further analysis substantiates that our model can exhibit stable performance in low-resource environments.

**Keywords:** ABSA · BERT · Prompt learning

## 1 Introduction

Aspect-based sentiment analysis(ABSA) is a branch of sentiment analysis, which aims to extract all the aspects and their corresponding sentiments within the sentence simultaneously [1]. Recent ABSA studies concentrated on three sub-tasks, i.e., Aspect Sentiment Classification(ASC) [2], Aspect Term Extraction (ATE) [12], Aspect Sentiment Triplet Extraction (ASTE) [3]. ASC determines the sentiment polarity of given aspects in a sentence. For example, given the sentence “The food in this restaurant is very good, but the service is bad”. This sentence mentions two aspects: Food and service, and for the ASC task, the purpose is to give the sentiment polarity of the two aspects “food” and “service” as positive and negative, respectively.

Recently, The emergence of large-scale pre-trained language models, such as Bidirectional Encoder Representations from Transformers (BERT) [4], has ushered natural language processing into a new era. Through training on a large corpus of Wikipedia documents and books, BERT acquires a nuanced understanding of language, syntax, and semantics through contextual analysis. In ABSA, recent work [10, 11, 15, 21] achieved appealing results based on pre-training models with BERT. Although great success has been achieved by the above studies, some critical problems remain when directly applying attention mechanisms or fine-tuning the pre-trained BERT in the task of ABSA.

Specifically, when BERT is used for downstream task fine-tuning, the [MASK] token does not appear, it only appears in the pre-training task. This creates a mismatch between pretraining and fine-tuning. Without the [MASK] token for fine-tuning, the model seems to have no starting point and no idea where to start. Meanwhile, simply initializing the encoder with a pre-trained BERT does not effectively handle informal expressions and complexity in ABSA as we expected. Thus, to better solve the above two types of problems, we propose a novel model, Prompt-oriented Fine-tuning Dual-BERT.

For the first challenge, we construct a Prompt-based BERT (ProBERT) by using sentence pair input with aspect words and a MASK token, which is used to indicate the emotional polarity of aspect words. This method is inspired by the idea of prompt learning because prompt-based fine-tuning with the objective of language modeling enables models to achieve significantly better performance on in-distribution cases than PLMs [19]. For the second, we construct a semantic aspect-based BERT (SemBERT) by utilizing a new attention mechanism, which combines the self-attention mechanism with the aspect-attention mechanism [23]. we expect that SemBERT could learn semantic representations different from syntactic representations. Our model concatenates the above two modules for emotional polarity classification and obtained good results on public datasets.

Our main contributions are as follows:

- (1) To the best of my knowledge, this is the first work to apply the idea of cue learning to ABSA. Moreover, it shows superiority in low-resource environments and Few-Shot Learning.
- (2) We propose a framework, Prompt-oriented Fine-tuning Dual-BERT, to improve the mismatch between pre-training and fine-tuning and the difficulty in analyzing the complex semantics of whole sentences and the syntactic semantics of aspects, respectively.
- (3) We conduct extensive experiments on three widely-used datasets, and the results demonstrate the effectiveness, rationality, and interpretability of the proposed model. Additionally, the source code and preprocessed datasets used in our work are provided on GitHub<sup>1</sup>.

<sup>1</sup> [https://github.com/yinwen2019/DualBERT\\_ABSA](https://github.com/yinwen2019/DualBERT_ABSA).

## 2 Related Work

### 2.1 ABSA

Aspect-based sentiment analysis (ABSA) is a task in sentiment analysis that aims to determine the sentiment polarity of a sentence in one or more specific aspects. ABSA allows for a deeper understanding of the sentiment expressed in a sentence by identifying the specific aspects that are being evaluated.

In recent years, graph neural networks have achieved widespread success on ABSA tasks, and graph neural networks are favored for their ability to better handle tree structures and resolve long-distance dependencies. [18] defined a new dependency tree structure based on the target aspect so that it is rooted in the target aspect and only the edges that have direct dependencies with the aspect are retained. To consider the type of dependency, [16] proposed a method based on a graph convolutional network that can use an attention mechanism to distinguish the importance of different edges, and proposes an attentive layer ensemble to learn from different levels of models. Although graph neural networks have shown certain advantages in ABSA tasks, there are still some drawbacks and limitations. Models based on graph structures are limited by the size of the graph, and their scalability is relatively poor.

The advent of pre-trained language models has significantly improved the accuracy and efficiency of aspect-based sentiment analysis. For instance, [20] proposed a model BERT4GCN, which integrates grammatical sequential features from BERT and syntactic structure information from the dependency graph to improve sentiment analysis tasks. Another study by [11] proposes two modules, parallel aggregation, and hierarchical aggregation, which improve the performance of BERT in extracting aspects and predicting the sentiment associated. To address the challenges caused by adding dynamic semantic changes to the ABSA task, [22] proposed a dynamic re-weighting BERT model (DR-BERT), which adds aspect-aware dynamic semantics to the learning framework of the pre-trained model. How to effectively leverage semantic understanding from BERT models remains a major challenge in current ABSA tasks.

### 2.2 Prompt Learning

A major existing problem in NLP is the need for task-specific supervised data, however, for many tasks, there are gaps in the amount of supervised data available. Prompt-based NLP learning methods try to solve this problem by learning a language model that is usually pre-trained first, to reduce or avoid the need for the number of supervised datasets. Most PLMS are pre-trained with modeling language objectives, whereas downstream tasks have very different objectives. To overcome the gap between pretraining and downstream tasks, some scholars [9, 13] have proposed to introduce prompt-tuning.

In the prompt-tuning paradigm, the downstream task is formalized as a language modeling problem by inserting a prompt template, and the results of language modeling can correspond to the solutions of downstream tasks. According

to [13], a prompt function consists of two parts: 1) Design a template with two slots: one slot [input] for inputting text  $x$  and another slot [MASK] for an intermediate generated answer text mask, which will be mapped to output  $y$ . 2) Fill the slot [input] with the input text  $x$ .

In general, for tasks related to generation, or solved using the standard autoregressive language model, prefix prompts tend to be more conducive to solving because they are well integrated with the left-to-right nature of the model. [9] proposed a prompt-based sentence embedding method, which enables BERT to achieve sentence embedding better. [8] pre-trained the prompt by adding soft prompts in the pre-training for the few-shot learning, to obtain better initialization. [5] proposed a modular framework called OpenPrompt, whose composability allows the freedom to combine different PLM, task formats, and prompt modules within a uniform paradigm. However, we find it a challenge to apply the generative hinting paradigm to ABSA in a high-quality way. Thus, our model utilizes prompt learning as an aspect of emotional representation and is the first to apply the prompts method to this problem.

### 3 Methodology

In this section, we will introduce the technical details of PFDualBERT. Specifically, we will begin with the problem definition, and then present an overview of the PFDualBERT architecture, which is illustrated by Fig. 1.

#### 3.1 Overview

**Problem Statement.** For Aspect Sentiment Classification (ASC), a sentence and a predefined aspect set  $(\mathbf{S}, \mathbf{A})$  is given. In this paper, we let  $\mathbf{S} = \{w_1, w_2, \dots, w_n\}$  and  $\mathbf{A} = \{a_1, a_2, \dots, a_m\}$  represent a sentence and a predefined aspect set, where  $n$  and  $m$  are the numbers of words in  $\mathbf{S}$  and the number of aspects in  $\mathbf{A}$ , respectively. For each  $\mathbf{S}$ ,  $\mathbf{A}_s = \{a_i | a_i \in \mathbf{A}, a_i \in \mathbf{S}\}$  denotes the aspects contained in  $\mathbf{S}$ . We treat each multiple-word aspect as a single word for simplicity, so  $a_i$  also means the  $i$ -th word of  $\mathbf{S}$ . The goal of ASC is to predict the sentiment polarity  $y_i \in \{\text{positive, negative, neutral}\}$  of the given aspect  $a_i \in \mathbf{A}_s$  in the input sentence.

**Overall Architecture.** As shown in Fig. 1, our proposed PFDualBERT takes the sentence and one of the aspects that appear in the text as the input and outputs the sentiment predictions of the aspects. It consists of two modules (i.e., Prompt-based BERT and Semantic aspect-based BERT), which share the same embedding input. 1) The Prompt-based BERT learns the word embedding output of the [MASK] token and fine-tunes the model using the Masked Language Model loss. 2) The Semantic-based BERT learns output representations associated with aspectual words through self-attention and aspect-aware attention mechanisms. 3) The sentiment classifier takes output representations of the above two modules to make predictions.

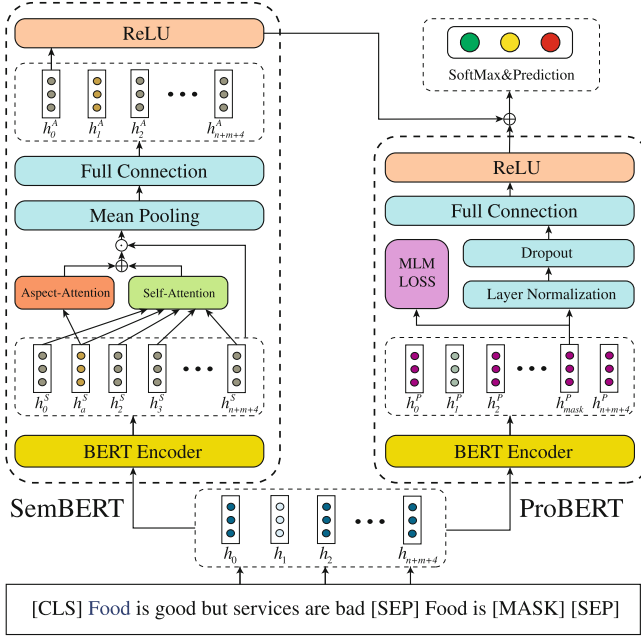


Fig. 1. Overall architecture of PFDualBERT

### 3.2 Context Encoder

To better represent the semantic information of aspect words and context words, we begin by mapping each word into a low-dimensional vector. For the input sequence, we construct an original sentence pair input, inspired by the idea of prompt learning. Specifically, the sentence pair for sequence input consists of the following:

$$[\text{CLS}] \textit{ sentence} [\text{SEP}] \textit{ aspect is} [\text{MASK}] [\text{SEP}] \quad (1)$$

where *sentence* and *aspect* represent whole-sentence input  $\mathbf{S}$  and a single aspect input  $a_i \in \mathbf{A}_s$ , respectively, [MASK] token represents the masked label in the BERT model. We use BERT to get our sentence embedding  $H = \{h_1, h_2, \dots, h_{n+m+4}\}$ .

In this paper, we leverage BERT [4] as a context encoder to extract hidden contextual representations. We get the output hidden states of the BERT encoder layer  $H^S = \{h_1^S, h_2^S, \dots, h_{n+m+4}^S\}$  and  $H^P = \{h_1^P, h_2^P, \dots, h_{n+m+4}^P\}$ , which are input into the ProBERT and SembERT modules respectively. Notably, the size of the embedded layer generated by the encoder used by the two modules is different. Specifically, the size of the hidden layer of the former is set to the BERT vocabulary size, which is used to prompt knowledge learning. In contrast, the size of the hidden layer of the latter is set as described in [4]. In the following section, we provide detailed information about our proposed PFDualBERT model.

### 3.3 Prompt-Based BERT (ProBERT)

The ProBERT module draws inspiration from prompt learning. Prior research has shown that prompt learning can effectively bridge the gap between pre-training and model tuning. In particular, this approach is highly effective in low-data regimes. These findings suggest that prompts can be used to more efficiently and effectively uncover the knowledge embedded within pre-trained language models, thereby leading to a deeper understanding of the underlying principles of these models. Based on these insights, we propose a Prompt-based BERT architecture that is specifically designed to extract important prompt knowledge learned during pre-training.

Specifically, We first take the sentence pair input and learn the overall semantics by the ProBERT encoder. Then, according to the semantics suggested by the [MASK] token, we construct an extra loss function to fine-tune the semantic association of aspect words. Finally, we input the learning representation of the [MASK] token into the classifier for polarity classification.

As illustrated in Fig. 1, our model selects the most important [MASK] token based on the overall prompt knowledge of the entire sentence, which then serves as the operative presentation information. Additionally, the ProBERT module takes the final outputs of the BERT encoder (i.e.  $\{h_i^P \in H^P\}$ ), where the hidden size is equal to the vocabulary size) as inputs. Then, we select the word representation embedding of the [MASK] token (i.e.  $h_{mask}$ , where  $mask$  represents the index of the [MASK] token in the sentence) as the next layer input, which represents the predicted information for the masked words.

After extracting the [MASK] token representation in the auxiliary clause of each sentence, we feed it into a Multilayer Perceptron (MLP) and map it to lower dimensions via a ReLU layer:

$$\begin{aligned} h_{mask}^* &= \text{LayerNorm}(h_{mask}^P) \\ d_{mask} &= \text{Dropout}(h_{mask}^*) \\ R_{mask} &= \text{Relu}(W_p d_{mask} + b_p) \end{aligned} \quad (2)$$

where  $W_p, b_p$  are learned parameters, while  $h_{mask}$  represents the MASK token representation selected from  $H^P$ .

**Masked Language Model Loss.** Furthermore, we utilize a special mask language model loss (MLM Loss) to optimize this module. Before doing so, we refactor the ground-truth label that is used to calculate our MLM Loss. Specifically, we assign three sentiment polarity words (i.e., *Positive*, *Negative*, and *Neutral*) based on the ground-truth label, which serves as the ground truth of prompt knowledge. Subsequently, we select the corresponding indices of sentiment words in the vocabulary and generate one-hot vectors, which serve as the ground truth word distribution for the [MASK] token prompt knowledge. Finally, we calculate the cross-entropy loss between the word distribution vector and the embedding of the [MASK] token in the final output of ProBERT, which

serves as the ultimate MLM loss:

$$\mathcal{L}_{mask} = - \sum_{(s,a) \in \mathcal{D}} \sum_{c \in \mathcal{V}} \log p(a) \quad (3)$$

where  $\mathcal{V}$  represents the set of all words in the vocabulary table and. But we're really only looking at labels that relate to emotional polarity. In other words, the loss function is calculated by the distribution probability of the corresponding words determined according to the polarity of the target. We believe that this loss function can effectively optimize the intermediate layers of ProBERT.

### 3.4 Semantic-Based BERT (SemBERT)

In contrast to ProBERT, SemBERT does not utilize [MASK] prompt knowledge. Instead, SemBERT obtains an aspect-based attention matrix in the form of an adjacency matrix through a self-attention and aspect-aware mechanism. The attention mechanism is a commonly used method for capturing interactions between aspect and context words [7]. To enhance the semantic features, we adopt the aspect-attention and self-attention mechanism as proposed in [23]. Specifically, we learn the attention scores from the output of the SemBERT Encoder. Next, we will provide a detailed introduction to these two mechanisms.

**Self-Attention.** Self-attention [17] is a technique that calculates the attention score of each pair of elements in parallel, allowing for the capture of interactions between any two words within a sentence. This involves computing a query and a key, which are then used to determine the attention score.

$$A_{self} = \frac{QW^Q \times (KW^K)^T}{\sqrt{d}} \quad (4)$$

where matrices  $Q$  and  $K$  are both equal to the word representations of the last layer of our SemBERT encoder  $H^S$ , while  $W^Q \in \mathbb{R}^{n \times n}$  and  $W^K \in \mathbb{R}^{n \times n}$  are both learnable weights. In addition,  $d$  is the dimensionality of the input node feature.

**Aspect-Attention.** ABSA task requires modeling the specific semantic correlation between the aspect term and its context sentence, which is different for each aspect term. To capture this correlation, we compute an additional attention score matrix  $A_{aspect}$  using aspect-attention [23]. The aspect-attention mechanism allows the model to attend more to the words that are related to the aspect term and ignore the irrelevant words in the sentence. This attention score matrix is computed by considering the aspect term as a query and the contextual words as keys, and it is used to weight the contextual word representations in the final classification layer.

$$A_{aspect} = \tanh \left( Q_a W^a \times (KW^K)^T + b \right) \quad (5)$$

where matrices  $K$  is equal to  $H^S$  produced by SemBERT encoder.  $W^a \in \mathbb{R}^{n \times n}$  and  $W^K \in \mathbb{R}^{n \times n}$  are both learnable weights. We compute aspect representation by applying mean pooling on  $h_a$  and copying it  $n$  times to obtain  $Q_a \in \mathbb{R}^{n \times n}$ . Then, we integrate the aspect-attention score with the self-attention score:

$$A = A_{self} + A_{aspect} \quad (6)$$

After obtaining the attention score matrix, we multiply it with the original word representation to obtain sentiment knowledge. Then, we use polarity classifiers to predict the sentiment of the aspect term in the given context sentence.

$$\begin{aligned} H^A &= A \times H \\ E_{cls} &= \text{Relu}(W_c(h_{cls}^A) + b_c) \end{aligned} \quad (7)$$

where  $W_c, b_c$  are learnable parameters, while  $h_{cls}^A$  represents the first token representation selected from  $H^A$ .

### 3.5 Model Training

After passing through ProBERT and SemBERT, the original word embeddings ( $H$ ) are respectively transformed into feature representations  $R$  and  $E$ . Finally, we utilize the Softmax function for sentiment polarity classification:

$$\hat{y} = \text{Softmax}(W_{sem}E_{cls} + W_{pro}R_{mask} + b) \quad (8)$$

where  $W_{sem}, W_{pro}, b$  are learnable parameters and bias, while  $\hat{y}$  is the predicted sentiment polarity distribution.

Finally, we apply the cross-entropy loss function for model training:

$$\mathcal{L} = - \sum_{(s,a) \in \mathcal{D}} \sum_{c \in \mathcal{C}} \log p(a) \quad (9)$$

where  $\mathcal{D}$  contains all the sentence-aspect pairs and  $a$  represents the aspect appearing in sentence  $s$ .  $\theta$  represents all the trainable parameters and  $\mathcal{C}$  is the collection of sentiment polarities.

Finally, we add the model loss and the MLM loss mentioned above to obtain the total loss. Our training goal is to minimize the following total objective function:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L} + \lambda_2 \mathcal{L}_{mask} + \beta \|\theta\|_2^2 \quad (10)$$

where  $\lambda_1, \lambda_2$  and  $\beta$  are regularization coefficients and  $\theta$  represents all trainable model parameters.

## 4 Experiment

### 4.1 Datasets

The experiments were conducted on three benchmark ABSA datasets: SemEval 2014 Task 4 Restaurant and Laptop reviews [14], and Twitter posts [6]. Each data item was labeled with one of the three sentiment polarities: positive, negative, or neutral. The statistical information of the dataset is shown in Table 1.



**Table 1.** Statistics for the three experimental datasets.

Dataset	#positive		#negative		#neutral	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	727	807	196	637	196
Laptop	976	337	851	128	455	167
Twitter	1507	172	1528	169	3016	336

### 4.2 Implementation Details

For our experiments, we initialize word embeddings with the official bert-large-uncased<sup>2</sup> models provided by [4] ( $n_{layers}=24$ ,  $n_{heads}=16$ ,  $n_{hidden}=1024$ ). The learning rate is set as 2e-5 and the dropout rate is set as 0.3. The batch size is manually tested in [16]. The hyper-parameter  $\lambda_1$ ,  $\lambda_2$  and  $\beta$  have been carefully adjusted, and final values are set to 0.5, 0.5, and 100 respectively. The model is trained using the Adam optimizer and evaluated by two widely used metrics. We run our model three times with different seeds and report the average performance.

### 4.3 Baselines

To comprehensively evaluate the performance of our model, we compare it with state-of-the-art baselines:

a)BERT-PT [21] explore a novel post-training approach on BERT to enhance the performance. b)BERT-SPC [15] feeds sequence “[CLS] + context + [SEP] + target + [SEP]” into the basic BERT model for sentence pair classification task. c)AEN-BERT [15] proposes an Attentional Encoder Network. d)BERT-AT [10] proposes a novel architecture called BERT Adversarial Training to utilize adversarial training. e)R-GAT+BERT [18] proposes a relational graph attention network. f)BERT4GCN [20] integrates the grammatical sequential features from BERT and the syntactic knowledge from dependency graphs. g)TGCN-BERT [16] uses an attentive layer ensemble to learn the contextual information from different GCN layers. h)SSEGcn-BERT [23] proposes a novel Syntactic and Semantic Enhanced Graph Convolutional Network. i)DR-BERT [22] proposes a novel method designed to learn dynamic aspect-oriented semantics.

### 4.4 Main Results

To demonstrate the effectiveness of PFDualBERT, we compared our model with previous works using accuracy and F1-score as evaluation metrics. The results, as shown in Table 2, indicates that our PFDualBERT model outperforms all previous works on the three datasets. Our comparison of non-specific BERT models (i.e., BERT and BERT-PT) with task-specific models (e.g., DGEDT-BERT and

<sup>2</sup> <https://github.com/huggingface/transformers>.

**Table 2.** Experimental results comparison on three publicly available datasets.

Models	Restarant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
BERT-PT [21]	84.95	76.96	78.07	75.06	–	–
BERT-SPC [15]	84.46	76.98	78.99	75.03	74.13	72.73
AEN-BERT [15]	83.12	73.76	79.93	76.31	74.71	73.13
BERT-AT [10]	86.03	79.24	79.35	76.50	–	–
RGAT-BERT [18]	86.60	81.35	78.21	74.70	76.15	74.88
BERT4GCN [20]	84.75	77.11	77.49	73.01	74.73	73.76
TGCN+BERT [16]	86.61	79.95	80.88	77.03	76.45	75.25
SSEGCN+BERT [23]	87.31	81.09	81.01	77.96	77.40	76.02
DR-BRET [22]	87.72	82.31	81.45	78.16	77.24	76.10
Our PFDualBERT	<b>88.11</b>	<b>82.62</b>	<b>83.07</b>	<b>80.03</b>	<b>77.84</b>	<b>76.73</b>

TGCN+BERT) for ABSA, revealed that task-specific BERT models perform better than non-specific models. Furthermore, we observed a performance trend where DR-BERT > SSEGCN+BERT > T-GCN > RGAT-BERT > AEN-BERT > BERT-PT. It can be inferred from this trend that aspect-related information is a critical factor influencing the performance of ABSA models. Despite the outstanding performance of previous models, ourPFDualBERT still outperforms the most advanced baseline (i.e., DR-BERT or SSEGCN-BERT) no matter in terms of Accuracy or F1-score. The results indicate that our strategy based on prompt knowledge and aspect attention is effective. At the same time, this also suggests that our proposed integration of two BERT together can better capture the deep semantics of sentences.

**Table 3.** Experimental results comparison on three publicly available datasets

Models	Restarant		Laptop		Twitter	
	Acc	F1	Acc	F1	Acc	F1
Our PFDualBERT	<b>88.11</b>	<b>82.62</b>	<b>83.07</b>	<b>80.03</b>	<b>77.84</b>	<b>76.73</b>
<i>w/o</i> self-attention	84.93	79.17	79.29	76.36	75.02	74.17
<i>w/o</i> aspect-attention	86.29	80.96	81.13	78.06	76.18	74.27
<i>w/o</i> mask language model loss	87.36	81.08	81.99	79.03	77.03	73.43

### 4.5 Ablation Study

As shown in Table 3, we conducted an ablation study to examine the effectiveness of different modules in PFDualBERT. We considered the basic PFDualBERT as the baseline model. The results reveal that removing the self-attention module significantly degrades the performance, confirming the necessity of the global semantics of the sentence for ABSA. We also observed that removing the aspect-attention module resulted in unsatisfactory performance, indicating that capturing aspect-related semantics is crucial, leading to a 1.82%, 1.94%, and 1.66% reduction in accuracy on the three datasets, respectively. This highlights the importance of aspect-attention in capturing correlated semantic information between aspects and contextual words. Additionally, removing the masked language loss module resulted in a performance drop of 0.75%, 1.08%, and 0.81% in accuracy on three datasets, respectively, indicating that the masked language model can assist in better learning prompt knowledge in the original encoder. Finally, removing masked language loss and aspect-attention led to a drop in performance, further emphasizing their crucial roles in PFDualBERT for ABSA. In conclusion, the ablation experimental results demonstrate that each component contributes significantly to the effectiveness of our entire model.

**Table 4.** Experimental results comparison on Laptop datasets by sampling datasets of various proportions.

Dataset	Train Size	25% Data		50% Data		75% Data		100% Data	
	Model	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Laptop	SSEGCN+BERT	76.18	72.55	78.32	74.06	79.91	75.65	81.01	77.96
	PFDualBERT	79.59	76.46	81.65	78.39	82.14	78.18	83.07	80.06
	<b>vs.</b>	+3.41	+3.91	+3.33	+3.79	+2.23	+2.53	+2.06	+2.10

### 4.6 Few-Shot Study

In order to investigate the effectiveness of PFDualBERT in low-resource settings, we conducted a few-shot study using various scaled versions of the Laptop dataset, with proportions of 25%, 50%, and 75%. SSEGCN [23] was used as the baseline model for comparison. As depicted in Fig. 4, the results show that PFDualBERT outperforms SSEGCN in terms of accuracy and F1-score. Moreover, as the size of the dataset decreases, both the accuracy and recall of the model decrease. However, PFDualBERT exhibits better performance than SSEGCN-BERT with smaller datasets, resulting in accuracy increases of 2.23%, 3.33%, and 3.41% for the 25%, 50%, and 75% proportions, respectively. These results suggest that PFDualBERT is more effective at handling sentiment information with limited training data in aspect-based sentiment analysis.

## 5 Conclusions

In this paper, we propose PFDualBERT, a novel model that integrates prompt knowledge and semantic correlations into aspect-based sentiment analysis. Our approach employs a unique sentence pair input that is based on prompts, which enhances the sentence encoding process. We evaluate PFDualBERT on three public datasets, and our experimental results demonstrate its effectiveness, especially in low-resource environments. Moreover, we plan to further optimize the number of model parameters and investigate other potential uses of our approach.

**Acknowledgement.** This research work was supported by the National Natural Science Foundation of China (NSFC) (U19A2059).

## References

1. Chen, W., Du, J., Zhang, Z., Zhuang, F., He, Z.: A hierarchical interactive network for joint span-based aspect-sentiment analysis. In: Proceedings of COLING 2022, pp. 7013–7019 (2022). <https://aclanthology.org/2022.coling-1.611>
2. Chen, X., et al.: Aspect sentiment classification with document-level sentiment preference modeling. In: Proceedings of ACL 2020, pp. 3667–3677 (2020)
3. Chen, Y., Chen, K., Sun, X., Zhang, Z.: A span-level bidirectional network for aspect sentiment triplet extraction. In: Proceedings of EMNLP 2022, pp. 4300–4309 (2022). <https://aclanthology.org/2022.emnlp-main.289>
4. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT 2019, pp. 4171–4186 (2019)
5. Ding, N., et al.: Openprompt: an open-source framework for prompt-learning. In: Proceedings of ACL 2022–System Demonstrations, pp. 105–113 (2022)
6. Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., Xu, K.: Adaptive recursive neural network for target-dependent twitter sentiment classification. In: Proceedings of ACL 2014, pp. 49–54 (2014)
7. Fan, F., Feng, Y., Zhao, D.: Multi-grained attention network for aspect-level sentiment classification. In: Riloff, E., Chiang, D., Hockenmaier, J., Tsujii, J. (eds.) Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3433–3442 (2018)
8. Gu, Y., Han, X., Liu, Z., Huang, M.: PPT: pre-trained prompt tuning for few-shot learning. In: Proceedings of ACL 2022, pp. 8410–8423 (2022)
9. Jiang, T., et al.: Promptbert: improving BERT sentence embeddings with prompts. In: Proceedings of EMNLP 2022, pp. 8826–8837 (2022). <https://aclanthology.org/2022.emnlp-main.603>
10. Karimi, A., Rossi, L., Prati, A.: Adversarial training for aspect-based sentiment analysis with BERT. In: ICPR 2020, pp. 8797–8803 (2020)
11. Karimi, A., Rossi, L., Prati, A.: Improving BERT performance for aspect-based sentiment analysis. In: 4th International Conference on Natural Language and Speech Processing, Trento, Italy, 12–13 November 2021, pp. 196–203 (2021)
12. Kumar, A., Srikanth, V.A., Narapareddy, V.T., Aruru, V., Neti, L.B.M., Malapati, A.: Aspect term extraction for opinion mining using a hierarchical self-attention network. *Neurocomputing* **465**, 195–204 (2021)

13. Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., Neubig, G.: Pre-train, prompt, and predict: a systematic survey of prompting methods in natural language processing. *ACM Comput. Surv.* **55**(9), 195:1–195:35 (2023)
14. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: aspect based sentiment analysis. In: *Proceedings of SemEval@COLING 2014*, pp. 27–35 (2014)
15. 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>
16. Tian, Y., Chen, G., Song, Y.: Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In: *Proceedings of NAACL-HLT 2021*, pp. 2910–2922 (2021)
17. Vaswani, A., et al.: Attention is all you need. In: *Advances in Annual Conference on Neural Information Processing Systems 2017*, pp. 5998–6008 (2017). <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
18. Wang, K., Shen, W., Yang, Y., Quan, X., Wang, R.: Relational graph attention network for aspect-based sentiment analysis. In: *Proceedings of ACL 2020*, pp. 3229–3238 (2020)
19. Wang, L., Lepage, Y.: Masked prompt learning for formal analogies beyond words. In: *Proceedings of (IJAI-ECAI 2022)*. *CEUR Workshop Proceedings*, vol. 3174, pp. 1–14 (2022). <http://ceur-ws.org/Vol-3174/paper1.pdf>
20. Xiao, Z., Wu, J., Chen, Q., Deng, C.: BERT4GCN: using BERT intermediate layers to augment GCN for aspect-based sentiment classification. In: *Proceedings of EMNLP 2021*, pp. 9193–9200 (2021)
21. Xu, H., Liu, B., Shu, L., Yu, P.S.: BERT post-training for review reading comprehension and aspect-based sentiment analysis. In: *Proceedings of NAACL-HLT 2019*, pp. 2324–2335 (2019)
22. Zhang, K., et al.: Incorporating dynamic semantics into pre-trained language model for aspect-based sentiment analysis. In: *Findings of ACL 2022*, pp. 3599–3610 (2022)
23. Zhang, Z., Zhou, Z., Wang, Y.: SSEGCN: syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In: *Proceedings of NAACL 2022*, pp. 4916–4925 (2022)