Prompt-Based Learning for Aspect-Level Sentiment Classification

Guowei Li, Fuqiang Lin, Wangqun Chen, Diwen Dong, and Bo Liu[✉]

National University of Defense Technology, Changsha 410073, China
{liguowei,linfuqiang13,chenwangqun19,ddw_bak,kyle.liu}@nudt.edu.cn

Abstract. The goal of Aspect-level Sentiment Classification (ASC) is to identify the sentiment polarity towards a specific aspect of a given sentence. Mainstream methods design complicated models and require a large scale of annotated training samples, and even perform finetuning based on pre-trained language models. Therefore, those supervised methods may be impractical in real-world scenarios due to the limited availability of labeled training corpora, a.k.a. low-resource settings. To this end, we propose an aspect-specific prompt learning approach (AS-Prompt) that fully utilizes the pre-trained knowledge and aspect-related information to deal with ASC tasks, enabling pre-trained models with huge parameters to achieve considerable results under the few-shot settings. Specifically, we transfer the sentiment classification task into Masked Language Modeling (MLM) by designing appropriate prompts and searching for the ideal expression of prompts in continuous space. Meanwhile, we integrate the prompts into the input sentence, thus adapting the model to the classification task under the guidance of sentiment labels. Experimental results on SemEval-2014 Task 4 show our proposed method achieves noticeably improvement compared with the original BERT models and discrete prompt methods. In addition, we test the performance of the model's transfer on different datasets and demonstrate the superiority of prompt learning when adapting to a new domain, especially under a low-resource setting.

Keywords: Aspect-level Sentiment Classification · Prompt Learning · Few-shot Learning

1 Introduction

As an essential part of the Aspect-Based Sentiment Analysis (ABSA) task, Aspect-Level Sentiment Classification (ASC) aims to predict the sentiment polarity of each given aspect term in a sentence [28]. This aspect-level sentiment analysis can provide more fine-grained guidance in making decisions. For example, suppose a restaurant knows specific reviews for various aspects of it. In this case, they can improve management by targeting niches to better meet the needs of consumers. From the comments in Fig. 1, we can infer that consumers rate the food positively but believe that restaurants should pay more attention to their service.
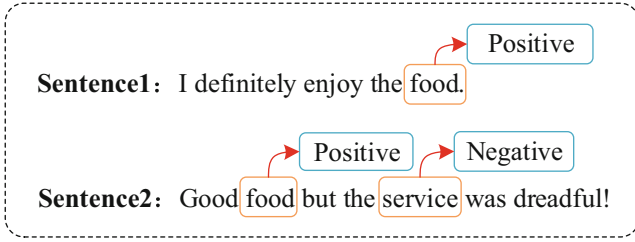
Fig. 1. Illustration examples of aspect-level sentiment classification. The aspect words and corresponding sentiment polarities are in orange and blue boxes, respectively. (Color figure online)

In recent years, researchers have proposed various neural network models for ASC tasks, which achieve promising performances [12,25]. Notably, pre-trained language models (PLMs) have been widely applied to ASC tasks considering the versatile knowledge contained in PLMs. A typical paradigm, called Fine-tuning, is to adapt PLMs to ASC tasks via additional parameters or well-designed objective function, and then optimize the whole model's parameters based on extensive task-specific training data. For example, Devlin and Sun et al. [5,19] construct a classifier to realize the ASC task based on the output of the pre-trained BERT. Nevertheless, the finetuning of such models relies on large-scale annotated corpora to ensure satisfied results, which is not easy to acquire due to costly labor. Besides, such methods suffer a marked decline in performance when transferring into a new domain [9].

Recently, researchers have realized the inefficiency of Fine-tuning under a low-resource setting, and propose a new paradigm named prompt-based learning. Instead of adapting the pre-trained LMs to the downstream tasks through target engineering, prompt-based methods reconstruct the downstream task to fit the LMs with the help of text prompts. Discrete prompts (e.g., The {aspect} made me feel [MASK]) are manually designed and bring significant improvements over the non-prompt methods [22]. Further, researchers automatically search discrete prompts and demonstrate the effectiveness of their works [11,23]. Their efforts in probing appropriate promots show that prompt-based learning methods bring considerable improvement in performance, especially under cross-domain and low-resource settings. However, those discrete prompt-based learning methods require a manual design of prompts, which is time-consuming and can only be optimized in discrete word embedding space.

To address these issues, we propose an aspect-specific prompt learning model (AS-Prompt) that automatically searches prompts with the guidance of aspect-specific information. Specifically, we design a soft prompt to adapt to the downstream ASC task by formulating the prompt as multiple learnable vectors and searching for a better representation of the prompt in continuous space. Such a way is also in line with the continuity feature of neural networks. Additionally, we leverage the sentiment label and aspect information to guide the training process by means of inserting the aspect words into the soft prompt, thus building

connections between aspect and corresponding sentiment. Finally, two types of MLM tasks are utilized to optimize the pre-trained model and the soft prompt with the reformulated input by concatenating the original sentence with the prompt.

Our main contributions can be summarized as follows:

– We propose an aspect-specific prompt method (AS-Prompt) to model the sentiment classification task into an MLM task, fully utilizing the aspect and sentiment label information to adapt the model to the downstream task.
– We demonstrate the ability of the model when transferring to a new domain, and explore the impacts of various factors on prompt-based methods to provide guidance for designing and training prompts.
– Experimental results on two datasets show that our method outperforms the baselines, especially under 16-shot and 64-shot scenarios.

2   Related Work

2.1   Aspect-Level Sentiment Analysis

Unlike the traditional coarse-grained sentiment classification task, fine-grained sentiment analysis has more practical values. Hu and Liu [10] is one of the early works to analyze aspect-related sentiment within the text. Pontik [18] contributes to the benchmark datasets of customer reviews for the research of aspect-based sentiment analysis (ABSA), which contain the restaurants and laptops domains. Research like Xu et al. [27] uses pre-trained language models to achieve significant results in sentiment classification, but those models are highly dependent on large-scale training datasets.

Since manually annotating the labels of text can be time-consuming, recent developments for ABSA attempt to finetune the models with unlabeled data. Sun et al. [24] finetune the BERT using a range of domain and task-related knowledge to compromise the effectiveness of a small dataset. Beigi et al. [1] propose an approach for sentiment analysis in unknown domains by adapting sentiment information with a generated domain-specific sentiment lexicon.

Apart from the performance on the small-scale training datasets, the ability to deal with multiple domains is another concern. Beigi and Moattaar [4] propose a transfer learning framework, which uses an adaptive domain model to eliminate the difference between domains. Cao et al. [30] adopt parameter transferring and attention sharing mechanisms to establish the connection between the source domain network and the target domain network. Zhao et al. [14] utilize a sentiment analyzer that learns sentiments via domain adaptive knowledge transfer to improve the classification performance. This paper also focuses on exploring effective approaches to improve the cross-domain ASC under a few-shot setting.

2.2   Prompt-Based Learning

Prompt-based learning reformulates the downstream tasks to adapt to the original LM training task. It utilizes the pre-trained LM to predict the desired output with appropriate prompts, even without additional task-specific training [6,8]. The prompt-based learning was applied to various domains as soon as it was proposed. For example, Yin et al. [29] and Schick et al. [20] explore prompt templates in classification-based tasks, where prompts can be easily constructed. The prompt-based learning reduces or eliminates the need for large supervised data sets for training models and can be applied to few-shot or zero-shot scenarios.

Prompt templates were first created manually according to human introspection [3,21]. But manual template engineering usually fails to find optimal prompts even with rich experience [11]. To find a way that allows LM to perform a task effectively, researchers examine continuous prompts in the embedding space of the model. Li and Liang [13] prepend a sequence of continuous task-specific vectors to the input while keeping the LM parameters frozen. Hambardzumyan et al. [7] propose to initialize the search for a continuous template with discrete prompts. Jiang et al. [11] propose "P-tuning" where continuous prompts are learned by inserting trainable variables into the embedded input.

Continuous prompts formulate the prompt as additional trainable parameters and search for an appropriate prompt by gradient optimization in embedding space. Some prompt-based learning models only update the parameters in prompts during the training stage, representative examples are Prefix-Tuning [13] and WARP [7]. Different from them, Liu and David et al. [2,15] optimize both parameters in pre-trained models and prompts, which is effective especially under few-shot settings. Our work introduces soft prompts to ASC tasks and utilizes aspect-specific information and sentiment label information to improve the performance of classification.

3   Methods

This section presents the task definition and the implementation of our proposed AS-prompt model. Unlike the discrete prompts, we follow the intuition that an appropriate prompt should be explored in continuous space. The architecture of the model is shown in Fig. 2. Our method adopts the pre-trained BERT as the pre-trained model. We set the tokens in prompts as trainable parameters and automatically optimize the prompts. The input of the method is reformulated by an original sentence and designed prompt.

3.1   Overview

We transfer the ASC task to the MLM task to avoid complicated training on pre-trained models based on prompt learning. Different from optimizing the model in discrete word embedding space, we adopt trainable vectors as prompt and finetune it in continuous space. The trainable prompt is more expressive than
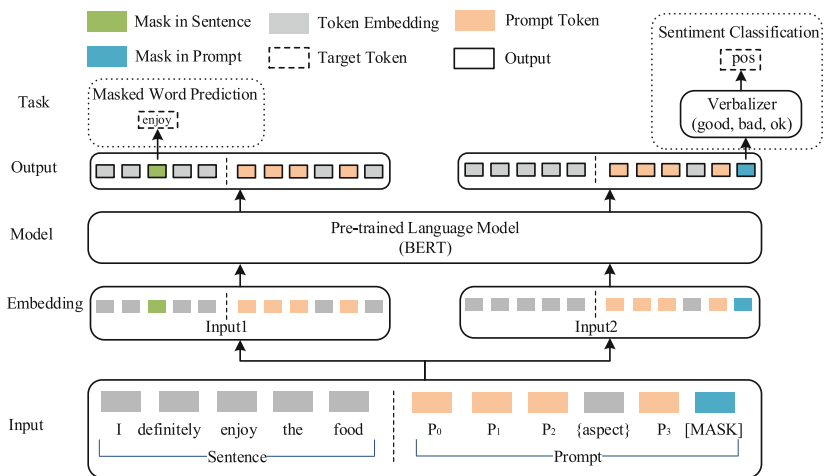
Fig. 2. The overall architecture of our model. The model design a continuous prompt to match the sentiment classification task and takes the original MLM as an auxiliary task. The {·} is the placeholder of the corresponding word. The prompt tokens are trainable vectors, while the embeddings of general tokens are fixed.

a discrete prompt formed of fixed words. To overcome the complex process of prompt selection, we directly initialize the continuous prompt with the same shape as the discrete prompt. Meanwhile, only finetuning the parameters in the prompt is far from enough for the model to learn domain-specific information. To this end, we implement the traditional MLM task of BERT as an auxiliary task to provide more semantic guidance for the model. Since the computation of training can be omitted under few-shot settings, we finetune both the model and the prompt in this paper.

3.2   Task Formulation

For a given sentence $X = [x_1, x_2, \cdots, x_n]$, the ASC task aims to identify the corresponding sentiment polarity s of each aspect a contained in the sentence, where the $s \in \{pos, neg, neu\}$. Let $\mathcal{V}$ refers to the vocabulary space of a language model $\mathcal{M}$. A prompt template T is denoted by $T = \{[p_{0:i}], a, [p_{i+1:m}], y\}$, where $p_i \in \mathcal{V}$ refers to the ith prompt token, a is the aspect token and y is the [MASK] token. The main process of the architecture can be divided into two parts, the masked word prediction task to finetune the pre-trained model and the sentiment classification task to search appropriate prompt for ASC. The Fig. 3 shows the formulation of input for the mentioned two tasks.

Given the sentence and the prompt, we concatenate the sentence with the prompt as input, and formulate the original input into the *Input1* and *Input2* to get the embeddings as follows:
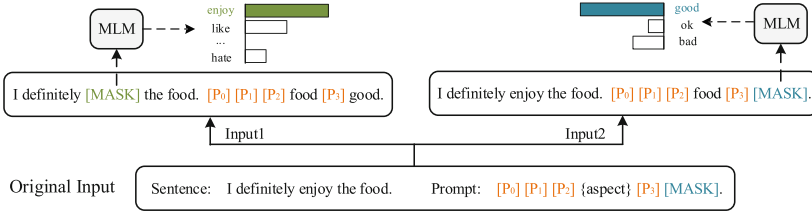
Fig. 3. An example of input formulation for the main processes of the model. $[P_i]$ in orange denotes the trainable prompt. [MASK] denotes the masked word and is differentiated by tasks. $\{\cdot\}$ is the placeholder of the corresponding aspect word. (Color figure online)

$$\{e(x_{0:j}), e(y_s), e(x_{j+1:n}), h_{0:i}, e(a), h_{i+1:m}, e(y)\}, \tag{1}$$

where $h_i (0 \le i < m)$ is a trainable vector and $e(x_i)$ is the initialization embedding of corresponding word.

For each input of MLM task, we compute the cross-entropy loss of the predictions on masked tokens. For the main task, the loss of *Input2* is:

$$\mathcal{L}_{prompt} = - \sum \sum y log(p(y|X, T)). \tag{2}$$

Similarly, we can get the loss of *Input1* as follows:

$$\mathcal{L}_{sen} = - \sum \sum y_s log(p(y_s|X, T)). \tag{3}$$

Then we can find a suitable prompt with the downstream loss function $\mathcal{L}$ by differentially optimize the prompt h_i:

$$\hat{h}_{0:m} = argmin_h(\mathcal{L}_{prompt} + \mathcal{L}_{sen}), \tag{4}$$

where the $\mathcal{L}_{prompt}$ refers to the loss of sentiment classification and $\mathcal{L}_{sen}$ refers to the loss of masked aspect prediction.

To better adapt the prompt, we convert the groudtruth labels $\{pos, neg, neu\}$ to $\{good, bad, ok\}$ and use the new labels for predicting.

3.3   In-Domain Data Pre-training

We can get available pre-trained weights for BERT, which has been trained on large corpora before. We prepare the pre-trained model with the extra in-domain datasets to include more domain-specific information. We adopt the same way as Seoh et al. [22], we only mask adjectives, proper nouns, and nouns, which are tightly related to sentiment. The baselines execute the same operation for a fair comparison. For the laptop domain, we use the reviews written for products from the electronics category in Amazon Review Data [17]. For the restaurants domain, we extract reviews related to restaurants from Yelp Open Dataset[1].

---

[1] https://www.yelp.com/dataset.

Table 1. SemEval 2014 dataset statistics after preprocessing.

Class	Laptops		Restaurants	
	Train	Test	Train	Test
Positive	987	341	2164	728
Negative	866	128	645	196
Neutral	460	169	496	196
Total	1850	638	3602	1120

4   Experiments

4.1   Datasets

We adopt SemEval 2014 Task 4 datasets released by Pontiki et al. [18] to measure the performance of our proposed model and baselines. It contains English review sentences from laptops and restaurants. The sentiment of each aspect is labeled as positive, negative, neutral, or conflict, where neutral refers to the opinion towards the aspect that is neither positive nor negative, and conflict denotes the existence of both positive and negative sentiment for an aspect. To conduct the experiments in the same condition as early studies [27], we remove the reviews labeled as conflict and split multiple aspect-sentiment labels within one text into different sentences. We select the training data from the training dataset by random numbers for each type of few-shot. The dataset statistics after preprocessing are shown in Table 1.

4.2   Baselines

We compare our proposed model with three Bert-based methods and two prompt-based methods:

– **BERT-ADA** [19] uses both domain-specific language model finetuning and supervised task-specific finetuning to realize the ASC task.
– **BERT [CLS]** [5] inserts a [CLS] token in front of the text and takes the corresponding output vector of [CLS] as the semantic representation of the text for classification.
– **BERT NSP** [5] aims to predict whether sentence B semantically follows sentence A when entering sentence A and sentence B simultaneously.
– **BERT LM** [22] transfers the ASC task as Language Modeling and designs discrete prompts as part of the input.
– **Null Prompts** [16] sets up a prompt template in the form of input text followed by [MASK] token for all tasks and automatically searches prompt in continuous space.

Table 2. Experimental results of our method and baselines. The best results are in bold. The underlined results indicate that our model's improvements over all baselines are statistically significant in the t-test of $p < 0.05$. The results of the first four baselines are taken directly from [22].

Laptops												
Model	4		16		64		256		1024		Full(1850)	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
BERT-ADA	-	-	-	-	-	-	-	-	-	-	79.19	74.18
BERT [CLS]	-	-	48.75	34.92	60.63	49.43	77.35	64.31	76.87	71.22	80.06	75.08
BERT NSP	-	-	48.24	31.35	60.91	49.27	72.38	64.64	76.77	71.12	80.25	75.46
BERT LM	61.60	51.29	69.05	58.60	72.80	65.54	76.59	70.65	**79.30**	**74.80**	**81.10**	**76.83**
Null Prompts	57.52	**54.15**	61.76	58.32	67.87	63.59	76.96	71.83	76.02	69.97	76.80	70.81
AS-Prompt	**62.23**	52.38	**71.79**	**64.47**	**75.39**	**69.77**	**77.74**	**72.98**	78.84	74.50	78.37	74.19
Restaurants												
Model	4		16		64		256		1024		Full(3602)	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
BERT-ADA	-	-	-	-	-	-	-	-	-	-	87.17	80.05
BERT [CLS]	-	-	58.89	34.50	73.00	50.79	79.45	64.70	83.48	73.62	86.77	79.33
BERT NSP	-	-	61.05	32.46	74.73	53.00	79.34	65.51	83.61	74.15	87.09	79.98
BERT LM	71.78	48.16	71.99	56.65	77.79	63.30	81.10	69.27	85.12	76.60	**87.50**	**80.78**
Null Prompts	63.93	48.88	74.00	**62.85**	76.43	61.92	79.20	68.83	83.72	73.52	83.58	77.18
AS-Prompt	**71.82**	**52.09**	**74.05**	60.73	**78.16**	**63.42**	**82.09**	**69.69**	**85.59**	**76.74**	85.64	76.71

4.3 Settings

We implement our model in PyTorch and load scripts for our datasets to be compatible with the Huggingface datasets[2]. We use spaCy for POS tagging and pytokenizations[3] for tokenizer alignment. All experiments are running on NVIDIA GeForce RTX 3090 GPU. For our MLM task, we utilize the pre-trained weights obtained from the transformers library [26]. The main layers of BERT are left frozen while training and do not get any updates. We only finetune the parameters in continuous prompts to cut down computation costs. We evaluate our model with randomly re-sampled training sets of size {4, 16, 64, 256, 1024, Full}, and the dataset is split following [19] under the setting of full-shot training. The training epoch is set to 20. As for results, we perform macro F1 score and accuracy (Acc.) as metrics to measure the performance of models. The initial learning rate is set to 0.00002 and we varied this value during training to reduce the training loss below 0.00001.

4.4 Overall Results

Table 2 records the overall results of the proposed model and baselines. As we can see from the table, all the prompt-based methods generally outperform the non-prompt ones in all few-shot cases for both target domains. It indicates that the

---

[2] https://huggingface.co/docs/datasets/master/.
[3] https://github.com/tamuhey/tokenizations.

Table 3. Accuracies of proposed model trained with in-domain and cross-domain data.

In/Cross	16		Full	
	Laptops	Restaurants	Laptops	Restaurants
In	71.79	74.05	78.37	85.64
Cross	73.48	71.70	79.29	76.33

prompt-based methods can easily formulate the downstream task to MLM task and fully take advantage of the knowledge contained in the pre-trained model. All prompt-based methods but BERT-LM achieve relatively lower performance than non-prompt methods under the full-shot setting, which suggests that finetining methods are superior to prompt-based methods when there are sufficient training data.

Considering the prompt-based methods, our proposed continuous prompt method achieves better results than the discrete prompt method in majority of scenarios. It implies that a fixed prompt template is not as powerful as a continuous prompt in building the connection between the pre-trained model and the downstream task. Our method achieves significant improvement, especially in a few-shot setting. With the decrease in training data, the performance of the benchmark methods decays significantly, while our approach remains at a high level. Considerable performance in few-shot indicates the effectiveness of the proposed model. Note that results on Restaurants overall higher than those on Laptops, we can speculate that the pre-trained model contains more knowledge in Restaurants domain than Laptops.

4.5   Further Analysis

To analyze the factors that may affect the prompt-based methods and provide guidance for prompt-learning methods, we conduct four more experiments for further discussion.

Transfer Ability. To examine the transferability of the model, we train the model on the in-domain dataset and test it on the cross-domain dataset. As we can see from Table 3, our model achieves better results on cross-domain datasets for laptops dataset under both the 16-shot and the full-shot settings. Results suggest that the prompt-based method has a strong ability to adapt to a new domain with considerable performance.

In-Domain Data Pre-training. We explore the impact of in-domain information on our method. We retrain the model with the in-domain dataset (i.e., Amazon for laptops and Yelp for restaurants) and compare the performance with the original BERT. Results in Table 4 show that retrained BERT model achieves much better outcomings than the original BERT model under low-resource settings. Still, the gap between them on the full-shot dataset is neglectable. It suggests that the prompt-based method is more practical for the sizeable pre-trained model, which already contains sufficient knowledge of various domains.

Table 4. Model's performance using the original pre-trained weights (Original) and the weights further trained with domain-specific review texts (Amazon, Yelp).

Datasets	Corpora	4		16		64		256		1024		Full(1850)	
		Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Laptops	Original	55.80	46.33	64.11	59.65	68.59	61.19	75.08	68.49	77.59	72.40	77.16	72.24
	+Amazon	62.23	52.38	71.79	64.47	75.39	69.77	77.74	72.98	78.84	74.50	78.37	74.19
Restaurants	Original	68.26	44.54	73.21	59.36	77.07	61.09	79.55	68.26	84.88	74.47	84.64	76.17
	+Yelp	71.82	52.09	74.05	60.73	78.16	63.42	82.09	69.69	85.59	76.74	85.64	76.71

Table 5. The impact of finetuning of the pre-trained model on 16-shot and full shot settings.

Finetuning	16		Full	
	Laptops	Restaurants	Laptops	Restaurants
Prompt+Model	71.79	74.05	78.37	85.64
Prompt	62.23	59.73	71.47	77.73

Type of Finetuning. Our model finetunes the parameters in pre-trained BERT and continuous prompt simultaneously under the assumption that the training costs of few-shot data can be neglected. As is shown in Table 5, we further freeze the parameters in BERT and compare the results with ours. We can conclude that there is no need to adjust the parameters in the pre-trained model when the training data is limited. However, finetuning the whole parameter helps when enough training data is provided.

Impact of Aspect. Intuitively, a better-designed prompt can improve the performance to a great extent. The results in Table 6 verify our conjecture. Here, we replace the aspect word with 'things' in the prompt and compare the results with ours. Results show that well-designed prompt largely improves the performance of the model.

Table 6. The impact of aspect information in prompts on full-shot dataset.

Prompt	Laptops		Restaurants	
	Acc	F1	Acc	F1
Prompt(aspect)	78.37	74.19	85.64	76.71
Prompt(things)	74.16	70.49	78.80	67.95

5   Conclusion

In this paper, we model the ASC task as LM and test the performance of our aspect-specific prompt learning model under the few-shot and the full supervised settings. Results demonstrate that the prompt learning method can achieve considerable performance on few-shot data while reducing the training cost of large

pre-trained models. Additionally, we reveal that the prompt-based approach is more practical to transfer to a new domain, and sufficient domain-specific knowledge contained in pre-trained model greatly improves the model's performance under the few-shot setting. In future work, since the prompt learning method can easily adapt to classification tasks and extraction tasks, it is possible to find a unified model to solve all subtasks of aspect-based sentiment classification based on prompt learning.

References

1. Beigi, O.M., Moattar, M.H.: Automatic construction of domain-specific sentiment lexicon for unsupervised domain adaptation and sentiment classification. Knowl. Based Syst. **213**, 106423 (2021)
2. Ben-David, E., Oved, N., Reichart, R.: PADA: a prompt-based autoregressive approach for adaptation to unseen domains. arXiv preprint arXiv:2102.12206 (2021)
3. Brown, T.B., et al.: Language models are few-shot learners. In: Proceedings of NeurIPS, pp. 1877–1901 (2020)
4. Cao, Z., Zhou, Y., Yang, A., Peng, S.: Deep transfer learning mechanism for fine-grained cross-domain sentiment classification. Connect. Sci. **33**(4), 911–928 (2021)
5. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL, pp. 4171–4186 (2019)
6. Gao, T., Fisch, A., Chen, D.: Making pre-trained language models better few-shot learners. In: Proceedings of ACL, pp. 3816–3830 (2021)
7. Hambardzumyan, K., Khachatrian, H., May, J.: WARP: word-level adversarial reprogramming. In: Proceedings of ACL, pp. 4921–4933 (2021)
8. Heinzerling, B., Inui, K.: Language models as knowledge bases: on entity representations, storage capacity, and paraphrased queries. In: Proceedings of EACL, pp. 1772–1791 (2021)
9. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. In: Gurevych, I., Miyao, Y. (eds.) Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, 15–20 July 2018, Volume 1: Long Papers, pp. 328–339. Association for Computational Linguistics (2018)
10. Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of ACM SIGKDD, pp. 168–177 (2004)
11. Jiang, Z., Xu, F.F., Araki, J., Neubig, G.: How can we know what language models know. Trans. Assoc. Comput. Linguist. **8**, 423–438 (2020)
12. Li, L., Liu, Y., Zhou, A.: Hierarchical attention based position-aware network for aspect-level sentiment analysis. In: Proceedings of CoNLL, pp. 181–189 (2018)
13. Li, X.L., Liang, P.: Prefix-tuning: optimizing continuous prompts for generation. In: Proceedings of ACL, pp. 4582–4597 (2021)
14. Li, Z., Qin, Y., Liu, Z., Wang, W.: Powering comparative classification with sentiment analysis via domain adaptive knowledge transfer. arXiv preprint arXiv:2109.03819 (2021)
15. Liu, X., et al.: GPT understands, too. arXiv preprint arXiv:2103.10385 (2021)
16. Logan, R.L., IV., Balažević, I., Wallace, E., Petroni, F., Singh, S., Riedel, S.: Cutting down on prompts and parameters: simple few-shot learning with language models. arXiv preprint arXiv:2106.13353 (2021)

17. Ni, J., Li, J., McAuley, J.J.: Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In: Proceedings of EMNLP, pp. 188–197 (2019)
18. Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., Manandhar, S.: Semeval-2014 task 4: aspect based sentiment analysis. In: Proceedings of COLING, pp. 27–35 (2014)
19. Rietzler, A., Stabinger, S., Opitz, P., Engl, S.: Adapt or get left behind: domain adaptation through BERT language model finetuning for aspect-target sentiment classification. In: Proceedings of LREC, pp. 4933–4941 (2020)
20. Schick, T., Schütze, H.: Exploiting cloze-questions for few-shot text classification and natural language inference. In: Proceedings of EACL, pp. 255–269 (2021)
21. Schick, T., Schütze, H.: It's not just size that matters: small language models are also few-shot learners. In: Proceedings of NAACL, pp. 2339–2352 (2021)
22. Seoh, R., Birle, I., Tak, M., Chang, H., Pinette, B., Hough, A.: Open aspect target sentiment classification with natural language prompts. In: Proceedings of EMNLP, pp. 6311–6322 (2021)
23. Shin, T., Razeghi, Y., IV, R.L.L., Wallace, E., Singh, S.: AutoPrompt: eliciting knowledge from language models with automatically generated prompts. In: Proceedings of EMNLP, pp. 4222–4235 (2020)
24. Sun, C., Huang, L., Qiu, X.: Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In: Proceedings of NAACL, pp. 380–385 (2019)
25. Tang, H., Ji, D., Li, C., Zhou, Q.: Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In: Proceedings of ACL, pp. 6578–6588 (2020)
26. Wolf, T., et al.: HuggingFace's transformers: state-of-the-art natural language processing. arXiv preprint arXiv:1910.03771 (2019)
27. 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, pp. 2324–2335 (2019)
28. Yan, H., Dai, J., Ji, T., Qiu, X., Zhang, Z.: A unified generative framework for aspect-based sentiment analysis. In: Proceedings of ACL, pp. 2416–2429 (2021)
29. Yin, W., Hay, J., Roth, D.: Benchmarking zero-shot text classification: datasets, evaluation and entailment approach. In: Proceedings of EMNLP, pp. 3912–3921 (2019)
30. Zhao, C., Wang, S., Li, D., Liu, X., Yang, X., Liu, J.: Cross-domain sentiment classification via parameter transferring and attention sharing mechanism. Inf. Sci. **578**, 281–296 (2021)