PRIOR-BERT AND MULTI-TASK LEARNING FOR TARGET-ASPECT-SENTIMENT JOINT DETECTION

Cai Ke[1,2], *Qingyu Xiong*[1,2,⋆], *Chao Wu*[1,2], *Zikai Liao*[3], *Hualing Yi*[1]

1. School of Big Data and Software Engineering, Chongqing University, Chongqing, China
2. Key Laboratory of Dependable Service Computing in
Cyber Physical Society (Chongqing University), Ministry of Education, Chongqing, China
3. Xidian University, Xian, China

ABSTRACT

Aspect-Based Sentiment Analysis (ABSA) is a fine-grained sentiment analysis task and has become a significant task with real-world scenario value. The challenge of this task is how to generate an effective text representation and construct an end-to-end model that can simultaneously detect (target, aspect, sentiment) triples from a sentence. Besides, the existing models do not take the heavily unbalanced distribution of labels into account and also do not give enough consideration to long-distance dependence of targets and aspect-sentiment pairs. To overcome these challenges, we propose a novel end-to-end model named Prior-BERT and Multi-Task Learning (PBERT-MTL), which can detect all triples more efficiently. We evaluate our model on SemEval-2015 and SemEval-2016 datasets. Extensive results show the validity of our work in this paper. In addition, our model also achieves higher performance on a series of subtasks of target-aspect-sentiment detection. Code is available at https://github.com/CQUPT-CaiKe/PBERT-MTL.

Index Terms— Aspect-Based Sentiment Analysis, Target-Aspect-Sentiment, Multi-Task Learning, Joint Detection

1. INTRODUCTION

Recently, Aspect-Based Sentiment Analysis (ABSA) [1] is drawing more and more attention from Nature Language Process (NLP) [2] researchers, which aims to analyze the sentiment of the aspect corresponding to a specific target in a sentence. For this purpose, the three main elements of the ABSA task include targets, aspects and sentiments. For instance, given the sentence "The food is great but the environment is bad". This sentence contains two aspects, FOOD#QUALITY and RESTAURANT#GENERAL. The corresponding targets are "food" and "environment", and the sentiment expressed are positive and negative.

The early ABSA task derived multiple related tasks under the different requirements. Initially, there were only single-element detections, such as target detection (TD) [3, 4] and aspect detection (AD) [5, 6] tasks which needed to satisfy the target or aspect is known [7]. If not, it is meaningless to transform independent sentiment detection into document-level sentiment detection task [8]. Furthermore, a dual-element detection task is a simultaneous detection to detect the double elements like target-aspect detection (TAD), target-sentiment detection (TSD), and aspect-sentiment detection (ASD). But for all the above detections mentioned, the sentiment dependence on the aspect and target cannot be captured simultaneously. Hence, it is very essential to construct an end-to-end model that can address the target-aspect-sentiment detection (TASD) task and take the interdependencies among these three elements into account.

As far as we know, there are only two related studies [9, 10], which respectively use traditional semantic lexicons and pre-trained BERT model to detect (target, aspect, sentiment) triples. Though they can indeed tackle the TASD task and detect (target, aspect, sentiment) triples, their methods still exist the following problems. First, Brun et.al [9] design a pipeline model but not an end-to-end multi-task joint detection model to deal with ABSA. Besides, pipelining TD and ASD tasks will lead to the wrong superposition of them, resulting in the poor effect. Second, Wan et.al [10] propose a joint detection model for ABSA, which utilizes pre-trained language model to detect all possible triples and achieves state-of-the-art results on the above tasks. However, there are heavily unbalanced distribution of labels after they reformulate datasets, which causes the model learning biased towards dominant labels. Moreover, the dependence of targets and aspect-sentiment pairs cannot be resolved in a longer distance. Therefore, these problems have a great impact on the improvement of the model on the TASD task.

To solve these problems above, we propose a novel end-to-end multi-task model named **P**rior-**BERT** and **M**ulti-**T**ask **L**earning (PBERT-MTL) to detect all (target, aspect, sentiment) triples more efficiently. Numerous experiments on

Corresponding author: cquxqy@163.com. This work was supported by the Major Special Program of Chongqing Science Technology Commission (No. CSTC2019jscx-zdztzxX0031), Graduate Scientific Research and Innovation Foundation of Chongqing (No. CYB19072 and No. CYS21067).

978-1-6654-0540-9/22/$31.00 ©2022 IEEE

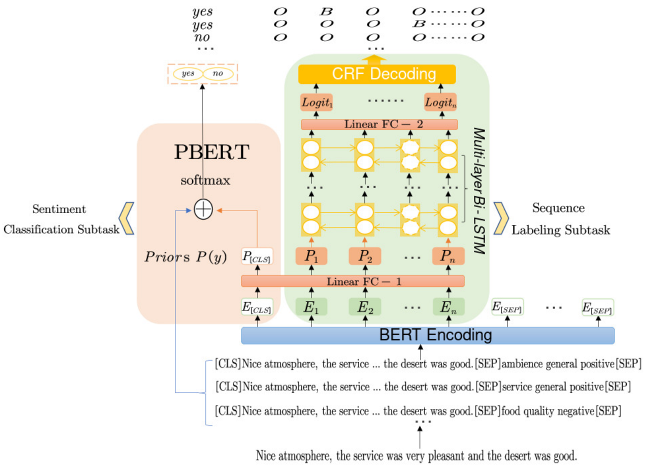

yes
yes
no
O B O O······O
O O O B······O
O O O O······O
yes no
CRF Decoding
Logit1
Logitn
Linear FC − 2
Multi-layer Bi-LSTM
PBERT
softmax
Sentiment Classification Subtask
Sequence Labeling Subtask
Priors P(y)
P[CLS]
P1
P2
Pn
Linear FC − 1
E[CLS]
E1
E2
En
E[SEP]
E[SEP]
BERT Encoding
[CLS]Nice atmosphere, the service ... the desert was good.[SEP]ambience general positive[SEP]
[CLS]Nice atmosphere, the service ... the desert was good.[SEP]service general positive[SEP]
[CLS]Nice atmosphere, the service ... the desert was good.[SEP]food quality negative[SEP]
Nice atmosphere, the service was very pleasant and the desert was good.

Fig. 1. The PBERT-MTL model overall architecture.

two standard aspect-level restaurant datasets SemEval-2015 Task 12 (denoted Res15) [11] and SemEval-2016 Task 5 (denoted Res16) [12] verify that our model performs well on the TASD task. Additionally, we also conduct experiments on its subtasks: TD, AD, ASD, TSD and TAD. The experimental results demonstrate that our PBERT-MTL model achieves higher performance on these tasks.

The **major contributions** of this paper are summarized as follows: (1) We propose Prior-BERT (PBERT) — a simple but universal method combining prior distribution knowledge of datasets with BERT for heavily unbalanced datasets. (2) We propose a novel end-to-end multi-task joint detection model (PBERT-MTL) to usefully address the challenges of the TASD task. (3) Extensive experiments have been carried out on two restaurant datasets to validate the proposed methods. The experimental results show that our model greatly improve the performance on the TASD task and its subtasks.

2. METHODOLOGY

The overall structure of PBERT-MTL model is shown in Fig.1. It is composed of the BERT encoder [13], the two fully-connected (FC) layers, the PBERT method for sentiment classification subtask, a multi-layer Bi-LSTM and a conditional random field (CRF) decoder [14] for sequence labeling subtask. In this section, we describe all modules of the model in detail.

2.1. Problem Definition

The TASD task, specifically, given by a sentence $S = \{s_1, s_2, \ldots, s_n\}$ consisting of n words, a predefined label set A as aspect categories and a predefined label set P as sentiment polarities, is to detect all possible triples (t, a, p) in S, where target t is a subsequence of S, a is an aspect category in A and p is a sentiment polarity in P.

2.2. Problem Reformulation

Inspired by [10, 15, 16, 17], we combine a sentence and an aspect-sentiment pair (a, p) as input to the model, namely every sentence S converted into $|A||P|$ sentences and aspect-sentiment pairs, where $|A|$ and $|P|$ are the number of sets A and P respectively. On this basis, we divide the TASD task into a sentiment classification subtask and a sequence labeling subtask. As shown in Fig.1, the former subtask is transformed into a binary classification problem, with "yes" meaning that there exists one target at least and "no" meaning that there exists no target. The latter subtask can be turned into a sequence labeling problem using the BIO labeling method, where "B" represents the starting word of a target, "I" an extended word of a target, "O" an unrelated word to a target. These two subtasks are closely related to each other and work at the same time, when given a S and a (a, p) pair, if the first subtask outputs "yes", there will be n triples (t, a, p) from the second subtask outputs where $n >= 0$. If $n = 0$, this special case means only one triple (NULL, a, p) with implicit target will be detected from S. Otherwise, If "no" is obtained from the first subtask, there will be no (t, a, p) triples detected.

2.3. Input and Embedding Layer

First, we construct an input sequence "[CLS] + a text sentence S + [SEP] + an aspect-sentiment pair (a, p) + [SEP]", where [CLS] and [SEP] are special labels introduced in the BERT model. The input sequence is composed of n+6 words, where the S contains n words and Fig.1 shows each (a, p) contains three words. This input sequence is then fed into the BERT encoder, outputting a d-dimensional embedding vectors at the final layer of the BERT encoder, they are $E_{[CLS]}$, E_{s1}, …, E_{sn}, $E_{[SEP]}$, …, $E_{[SEP]}$, where the two [SEP] inputs are transformed into two different vectors. After the first vector $E_{[CLS]}$ is fed into the FC-1 layer, the vector $P_{[CLS]}$ is defined as:

$$P_{[CLS]} = tanh(W_1 E_{[CLS]} + b_1) \tag{1}$$

where $W_1 \in \mathbb{R}^{d \times 2}$ and $b_1 \in \mathbb{R}^2$ are trainable parameters.

Second, the next n vectors $E_{s1}, \ldots, E_{sn}$ are fed into the same FC layer followed by a multi-layer Bi-LSTM and a CRF decoder to predict a label sequence in the BIO tagging scheme. More precisely, the vectors P_{si} computed form E_{si} (where $1 <= i <= n$) is defined as:

$$P_{si} = tanh(W_2 E_{si} + b_2) \tag{2}$$

where $W_2 \in \mathbb{R}^{d \times n}$ and $b_2 \in \mathbb{R}^n$ are trainable parameters.

2.4. The PBERT Method

Inspired by Menon et.al [18], we propose a simple method which combines the label priors $P(y)$ and the vector $P_{[CLS]}$ followed by a softmax decoder to effectively alleviate the heavy imbalance between "yes" and "no" labels shown in

Table 1. The details of Res15 and Res16. “yes” and “no” indicate the number of labels respectively.

Datasets		Aspects	Original			Reformulation			Implicit Targets
			sentences	yes	no	sentences	yes	no	
Res15	Train	13	1315	1	2	43642	1	38	375 (22.67%)
	Test	13	685	1	2	22660	1	38	248 (29.35%)
Res16	Train	12	2000	1	2	61453	1	35	627 (25.01%)
	Test	12	676	1	2	21097	1	35	208 (24.21%)

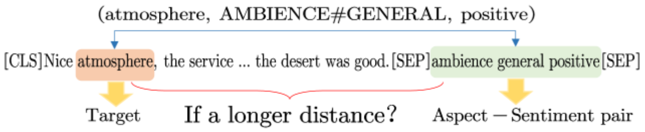

(atmosphere, AMBIENCE#GENERAL, positive)
[CLS]Nice atmosphere, the service ... the desert was good.[SEP]ambience general positive[SEP]
Target
If a longer distance?
Aspect − Sentiment pair

Fig. 2. The long-distance dependence between the target and aspect-sentiment pair.

Table 1. In detail, the probability distribution vector $g \in \mathbb{R}^2$ on the “yes/no” label is defined below:

$$g = softmax(P_{[CLS]} + \tau \cdot \log P(y_i)) \tag{3}$$

where τ is a tuning parameter to calibrate $P_{[CLS]}$ and y_i is the i^{th} element of $y \in \{yes, no\}$.

The loss value for predicting the “yes/no” label namely sentiment classification subtask is directly equivalent to:

$$loss_g = -\sum_{i=1}^{2} I(d(i) = label) \log(g_i) \tag{4}$$

where g_i is the i^{th} element of g, $d(1) = yes$, $d(2) = no$, and $I(x) = 1$ if x is true or $I(x) = 0$ if x is false.

Compared with the traditional softmax cross-entropy, PBERT adds the “yes/no” priors $P(y)$ from Table 1 to each $P_{[CLS]}$. In fact, not only their optimization purpose is the same [18], but also the increased prior knowledge can obtain further gains by combining with the vector $P_{[CLS]}$ which contains rich semantic features [13]. Moreover, the PBERT can readjust the weight between “yes” and “no” labels and sample more “yes” labels to help detect all possible targets.

2.5. Multi-layer Bi-LSTM and CRF Decoder

To address the long-distance dependence problem between targets and aspect-sentiment pairs considered in Fig.2. Motivated by Huang et.al [19], we employ a multi-layer Bi-LSTM to capture deep semantic information and the long-distance dependence. In detail, after the vectors $P_1, \ldots, P_n$ are fed into the Bi-LSTM, the hidden state $\overleftarrow{h} \in \mathbb{R}^{n \times d_{hid}}$ and $\overrightarrow{h} \in \mathbb{R}^{n \times d_{hid}}$ can be obtained from the forward and the backward LSTM, where d_{hid} is the number of the LSTM hidden units. The final feature representation $H = \{h_1, \ldots, h_n\} \in \mathbb{R}^{n \times 2d_{hid}}$ can be acquired by concatenating both LSTM hidden units.

After H are fed into the FC-2 layer, we then feed the outputs vectors $Logit_1, \ldots, Logit_n$ into the CRF model to consider the dependence relationship between adjacent labels and obtain the global optimal label sequence. In detail, the vectors $Logit_i$ (where $1 <= i <= n$) is defined as:

$$Logit_i = tanh(W_3 H + b_3) \tag{5}$$

where $W_3 \in \mathbb{R}^{2d_{hid} \times 3}$ and $b_3 \in \mathbb{R}^3$ are trainable parameters.

We denote $p(\mathbf{T}|\mathbf{P})$ by the probability for predicting a label sequence namely sequence labeling subtask, where $\mathbf{T}$ is a label sequence and $\mathbf{P}$ is a $n \times 3$ matrix composed of $Logit_1, \ldots, Logit_n$.

The loss value for predicting the label sequence $\mathbf{T} =< t_1, \ldots, t_n >$ is defined as:

$$loss_t = -\log(p(\mathbf{T}|\mathbf{P})) \tag{6}$$

2.6. Multi-task Learning

Our model follows the spirit of multi-task learning [20], the outputs will be obtained by optimizing the loss function of sentiment classification subtask and sequence labeling subtask at the same time. Hence, the multi-task loss is defined as follow:

$$loss = \sum_{i=1}^{N} ({loss_g}^i + {loss_t}^i) \tag{7}$$

where N is the number of all training tuples, and ${loss_g}^i$ and ${loss_t}^i$ represent the i^{th} training tuple of the two loss values.

3. EXPERIMENTS AND RESULTS

3.1. Experimental Settings

We evaluate PBERT-MTL model on SemEval-2015 Task 12 (denoted Res15) and SemEval-2016 Task 5 (denoted Res16). Table 1 shows the details of Res15 and Res16. For our model, micro-F1 score is used in percent as the evaluation criteria for TASD task and its subtasks.

To train our model, BERT is used by us to create the word embedding of the PBERT-MTL model. The number of layers and hidden units of Bi-LSTM are 5 and 128, respectively. The max sequence length is 128, the dropout probability is 0.1, and the τ is 1. We utilize Adam optimizer [27] to optimize gradient descent, whose learning rate is set at 2e-5 and the maximum epoch is set at 30.

Table 2. The predictions of our method and the predictions of the best existing method TAS-BERT are compared in two representative examples.

Text	Gold	Method	Prediction	Type

it was romantic - and even nice even with my sister, reminded me of italy, and had **artwork** and **music** that kept up the feeling of being in a Mediterrean villa.	{artwork, AMBIENCE#GENERAL, positive}
	{music, AMBIENCE#GENERAL, positive}
		TAS-BERT	{music,AMBIENCE#GENERAL,positive}	Error-1

		Our method	{artwork,AMBIENCE#GENERAL,positive}	Correct
			{music,AMBIENCE#GENERAL,positive}	Correct
the best **place** for a leisure sunday breakfast amidst yachts, then take a stroll through the nearby farmer's market.	{place, RESTAURANT #MISCELLANEOUS, positive}	TAS-BERT	{NULL,NULL,NULL}	Error-2
		Our method	{place,RESTAURANT#MISCELLANEOUS,positive}	Correct

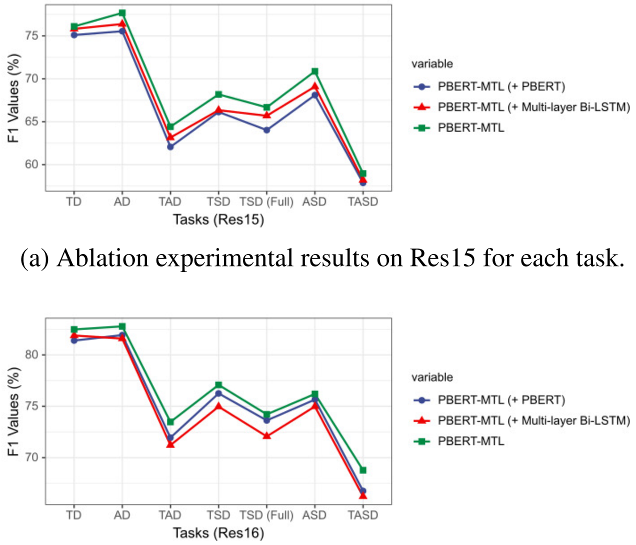
Table 3. Evaluation results of related tasks on Res15 and Res16. The bold means the average result of 10 random seeds. For the TSD task, scores outside "()" are for test sets without implicit targets, whereas scores in "()" are for the full test sets.

Tasks	Methods	Res15	Res16
TD	MTNA [21]	67.73	72.95
	DE-CNN [22]	-	74.37
	THA-STN [4]	71.46	73.61
	BERT-PT [23]	73.15	77.97
	TAS-BERT [10]	75.00	81.37
	PBERT-MTL	**75.66**±0.76	**82.16**±0.23
AD	BERT-pair-NLI-B [24]	70.78	80.25
	MTNA [21]	65.97	76.42
	TAN [6]	-	78.38
	Sentic LSTM+TA+SA [5]	73.82	-
	TAS-BERT [10]	76.34	81.57
	PBERT-MTL	**77.14**±0.55	**82.34**±0.47
TAD	TAS-BERT [10]	63.37	71.64
	PBERT-MTL	**64.21**±0.57	**72.97**±0.24
TSD	E2E-TBSA [25]	53.00	63.10
	DOER [26]	56.33	65.91
	TAS-BERT [10]	66.11 (64.29)	75.68 (72.92)
	PBERT-MTL	**67.53**±0.71(**66.12**)±0.67	**76.44**±0.14(**74.01**)±0.35
ASD	Baseline-1-f_lex [9]	-	63.50
	BERT-pair-NLI-B [24]	63.67	72.70
	TAS-BERT [10]	68.50	74.12
	PBERT-MTL	**70.43**±0.29	**75.88**±0.12
TASD	Baseline-1-f_lex [9]	-	38.10
	TAS-BERT [10]	57.51	65.89
	PBERT-MTL	**58.52**±0.23	**67.65**±0.34

3.2. Results and Analysis

In Table 3, our model obtains higher performance on the TASD task and its five subtasks. In addition, to verify the validity of the different key components (the PBERT and the multi-layer Bi-LSTM) of our model, we conduct ablation experiments to compare with the full PBERT-MTL model. These ablation experimental results are shown in Fig.3 and indicate the PBERT method and the multi-layer Bi-LSTM can further improve PBERT-MTL performance when they are combined by multi-task learning. In fact, even if one of the components is used, it outperforms the best model.

In Table 2, we give two representative examples to directly comprehend the main error case types and the different predictions between our model and the best model. The type Error-1 is the ignorance of the long-distance dependence between targets and aspect-sentiment pairs, and the type Error-2 is the ignorance of the heavily unbalanced labels distribution between "yes" and "no" labels. As for the first example, TAS-BERT cannot capture the longer distance dependence between the target "artwork" and its aspect-sentiment pair but our model captures this dependence. As for the second example, TAS-BERT cannot fully sample the "yes" labels, so that it detects no triple but our model detects the triple successfully.



(a) Ablation experimental results on Res15 for each task.

(b) Ablation experimental results on Res16 for each task.

Fig. 3. Ablation experiments of PBERT-MTL.

4. CONCLUSIONS

In this paper, we propose a novel end-to-end multi-task model named PBERT-MTL for TASD task which utilizes the proposed PBERT method to alleviate heavily unbalanced labels distribution and the multi-layer Bi-LSTM to capture the long-distance dependence. Experiments on SemEval-2015 and SemEval-2016 demonstrate that our model can detect (target, aspect, sentiment) triples efficiently and achieve higher performance on the TASD task and its subtasks.

5. REFERENCES

[1] Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou, "Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 6578–6588.

[2] Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco, "Affective computing and sentiment analysis," in *A practical guide to sentiment analysis*, pp. 1–10. Springer, 2017.

[3] Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao, "Coupled multi-layer attentions for co-extraction of aspect and opinion terms," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017, vol. 31.

[4] Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang, "Aspect term extraction with history attention and selective transformation," *arXiv preprint arXiv:1805.00760*, 2018.

[5] Yukun Ma, Haiyun Peng, and Erik Cambria, "Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive lstm," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, vol. 32.

[6] Sajad Movahedi, Erfan Ghadery, Heshaam Faili, and Azadeh Shakery, "Aspect category detection via topic-attention network," *arXiv preprint arXiv:1901.01183*, 2019.

[7] Yequan Wang, Aixin Sun, Minlie Huang, and Xiaoyan Zhu, "Aspect-level sentiment analysis using as-capsules," in *The World Wide Web Conference*, 2019, pp. 2033–2044.

[8] Maryem Rhanoui, Mounia Mikram, Siham Yousfi, and Soukaina Barzali, "A cnn-bilstm model for document-level sentiment analysis," *Machine Learning and Knowledge Extraction*, vol. 1, no. 3, pp. 832–847, 2019.

[9] Caroline Brun and Vassilina Nikoulina, "Aspect based sentiment analysis into the wild," in *Proceedings of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 2018, pp. 116–122.

[10] Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z Pan, "Target-aspect-sentiment joint detection for aspect-based sentiment analysis," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, pp. 9122–9129.

[11] Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis," in *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)*, 2015, pp. 486–495.

[12] Maria Pontiki, Dimitrios Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad Al-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al., "Semeval-2016 task 5: Aspect based sentiment analysis," in *International workshop on semantic evaluation*, 2016, pp. 19–30.

[13] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[14] Niklas Jakob and Iryna Gurevych, "Extracting opinion targets in a single and cross-domain setting with conditional random fields," in *Proceedings of the 2010 conference on empirical methods in natural language processing*, 2010, pp. 1035–1045.

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

[16] Xiao Liu, Kaixuan Ji, Yicheng Fu, Zhengxiao Du, Zhilin Yang, and Jie Tang, "P-tuning v2: Prompt tuning can be comparable to fine-tuning universally across scales and tasks," *arXiv preprint arXiv:2110.07602*, 2021.

[17] Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough, "Open aspect target sentiment classification with natural language prompts," *arXiv preprint arXiv:2109.03685*, 2021.

[18] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar, "Long-tail learning via logit adjustment," *arXiv preprint arXiv:2007.07314*, 2020.

[19] Zhiheng Huang, Wei Xu, and Kai Yu, "Bidirectional lstm-crf models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.

[20] Rich Caruana, "Multitask learning," *Machine learning*, vol. 28, no. 1, pp. 41–75, 1997.

[21] Wei Xue, Wubai Zhou, Tao Li, and Qing Wang, "Mtna: a neural multi-task model for aspect category classification and aspect term extraction on restaurant reviews," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2017, pp. 151–156.

[22] Hu Xu, Bing Liu, Lei Shu, and Philip S Yu, "Double embeddings and cnn-based sequence labeling for aspect extraction," *arXiv preprint arXiv:1805.04601*, 2018.

[23] Hu Xu, Bing Liu, Lei Shu, and Philip S Yu, "Bert post-training for review reading comprehension and aspect-based sentiment analysis," *arXiv preprint arXiv:1904.02232*, 2019.

[24] Chi Sun, Luyao Huang, and Xipeng Qiu, "Utilizing bert for aspect-based sentiment analysis via constructing auxiliary sentence," *arXiv preprint arXiv:1903.09588*, 2019.

[25] Xin Li, Lidong Bing, Piji Li, and Wai Lam, "A unified model for opinion target extraction and target sentiment prediction," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, vol. 33, pp. 6714–6721.

[26] Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang, "Doer: dual cross-shared rnn for aspect term-polarity co-extraction," *arXiv preprint arXiv:1906.01794*, 2019.

[27] Diederik P Kingma and Jimmy Ba, "Adam: A methodfor stochastic optimization," in *International Conference onLearning Representations (ICLR)*, 2015.