

Aspect-based Sentiment Analysis as Machine Reading Comprehension

Yifei Yang^{1,2}, Hai Zhao^{1,2,*}

¹Department of Computer Science and Engineering, Shanghai Jiao Tong University

²MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University

{yifeiyang}@sjtu.edu.cn, zhaohai@cs.sjtu.edu.cn

Abstract

Existing studies typically handle aspect-based sentiment analysis by stacking multiple neural modules, which inevitably result in severe error propagation. Instead, we propose a novel end-to-end framework, **MRCOOL: MRC-PrOmppt mOdeL** framework, where numerous sentiment aspects are elicited by a machine reading comprehension (MRC) model and their corresponding sentiment polarities are classified in a prompt learning way. Experiments show that our end-to-end framework consistently yields promising results on widely-used benchmark datasets which significantly outperform existing state-of-the-art models or achieve comparable performance.¹

1 Introduction

Compared with traditional sentence-level or document-level sentiment analysis tasks, aspect-based sentiment analysis (ABSA) requires finer grained analysis on the texts and extracting more detailed information (Liu, 2012; Pontiki et al., 2014a). ABSA contains many subtasks, such as aspect category detection, opinion term extraction (OE), etc. Aspect term extraction (AE) and aspect-level sentiment classification (SC) are two elemental subtasks of ABSA. AE means extracting the sentiment aspects from a given plain sentence and SC implies recognizing the sentiment polarities of the given aspects in a sentence. Combining the above two subtasks, aspect term extraction and sentiment classification (AESC) establishes the third fundamental subtask. The AE, SC and AESC for the sentence *Excellent food, although the interior could use some help.* are given in Figure 1.

In general, the existing mainstream approaches can be roughly divided into two brands. The first

Excellent food, although the interior could use some help.

AE	SC
<ul style="list-style-type: none">• <i>food</i>• <i>interior</i>	<ul style="list-style-type: none">• <i>food</i>-positive• <i>interior</i>-negative
Aspect Term Extraction and Sentiment Classification	
<ul style="list-style-type: none">• (<i>food</i>, positive)• (<i>interior</i>, negative)	

Figure 1: An example of AE, SC and AESC. AE needs to identify the aspect terms *food* and *interior*. When given these two terms, SC should recognize the sentiment polarities of them as positive and negative separately. AESC is going to complete these two tasks from the given sentence.

employs the two-stage method which first accomplishes the AE and is followed by another model to perform SC, thus achieving the AESC (Yu et al., 2018; Hu et al., 2019; Fan et al., 2019). The second tries to fulfill the three subtasks by a more unified methodology which extracts the terms and their corresponding polarities in a joint or interactive way (Liu et al., 2016; Wang et al., 2017; Li and Lam, 2017; Fan et al., 2018; He et al., 2019; Luo et al., 2019; Li et al., 2019; Peng et al., 2020; Chen and Qian, 2020; Chen et al., 2020; Wan et al., 2020). However, the above mentioned schemes perform ABSA task by stacking recurrent neural networks (RNN) or attention mechanisms and usually lead to too complex models.

In recent years, machine reading comprehension (Li et al., 2020; Liu et al., 2020; Su et al., 2020; Mao et al., 2021) quickly become a hot topic among various challenging natural language understanding tasks. Generally, MRC model may give a proper answer for a query based on a given passage. There are various types of MRC tasks according to the desired answer forms, among which span MRC or extractive MRC draw quite a lot of attention (Glass et al., 2019; Wu et al., 2019; Zhang et al., 2020). Most of the progress for MRC may be attributed to the latest pre-trained language models (PrLMs)

*Corresponding author. This work was supported by Key Projects of National Natural Science Foundation of China (U1836222 and 61733011).

¹<https://github.com/yangyifei729/MRC4absa>

The waiters were very professional, courteous and attentive.

<p>AE as MRC Input to PrLM (before embedded) [CLS] find the sentiment aspect terms in the text. [SEP] The waiters were very professional, courteous and attentive. Output of MRC: <i>waiters</i></p>
<p>SC as Prompt Learning Input to PrLM (before embedded) [CLS] The waiters were very professional, courteous and attentive. [SEP] I felt the <i>waiters</i> was [MASK]. [CLS] The waiters were very professional, courteous and attentive. [SEP] I [MASK] the <i>waiters</i>. [CLS] The waiters were very professional, courteous and attentive. [SEP] The <i>waiters</i> made me feel [MASK]. [CLS] The waiters were very professional, courteous and attentive. [SEP] The <i>waiters</i> is [MASK]. Output of Prompt Learning: <i>positive</i></p>

Figure 2: An example of modeling AE as MRC task and SC as prompt learning task. This figure shows the final input of two subtasks to the PrLM.

such as BERT (Devlin et al., 2019). For enhancing multiple downstream tasks including MRC, PrLM may serve as a powerful enough encoder in the corresponding model for effectively capturing salient features from input text (they are passage and query in MRC).

Prompt learning is a natural manner to leverage the knowledge of PrLM which requires adapting the downstream tasks into a self-supervised learning task of the corresponding PrLM. For example, Chen et al. (2021) convert the relation extraction task to the masked language model (MLM) task of BERT and Sun et al. (2021) apply the next sentence prediction (NSP) task to carry out the downstream tasks. Even though MRC and prompt learning can take advantage of the knowledge of PrLM and facilitate the performance on downstream tasks, these paradigms still have obvious defects, (1) The query for MRC can severely inhibit the performance of downstream tasks, but the construction of query is currently based on templates or empiric which leads to huge labor costs and it is not guaranteed to find the best matched query. (2) For the prompt learning, after getting the output of PrLM, a standardized process is applying a verbalizer to project the original labels to the label words of the downstream task. When adopting MLM to perform tagging tasks, the current verbalizer selects the probability distribution of a few specific words from the output word-embedding of [MASK] token to determine the final prediction. This manner makes the verbalizer very sparse and can not make full use of the knowledge of PrLM.

To alleviate the above issues, we propose a novel end-to-end framework named **MRCOOL** to handle the AE, SC and AESC once for all. For AE, we

model it as an MRC task and propose a query encoder to search for the possible latent optimal query in a continuous space. We treat SC as a prompt learning task and apply a concise MLP verbalizer to reduce the sparsity. Our experiments are conducted on three widely-used benchmark datasets. Results show that our framework outperforms the current methods or gets comparable performance.

2 MRCOOL Framework

2.1 Task Formulation

The three subtasks can be formulated as a tuple extraction task. Given an input sentence $X = \{x_1, x_2, \dots, x_n\}$ of length n , the corresponding output is $Y = \{(a_1, p_1), (a_2, p_2), \dots, (a_m, p_m)\}$ where a_i indicates an aspect and p_i represents its polarity.

Given a training dataset $\mathcal{D} = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_{|\mathcal{D}|}, Y_{|\mathcal{D}|})\}$, the purpose of our framework is to maximize the likelihood:

$$L(\mathcal{D}) = \prod_{i=1}^{|\mathcal{D}|} \prod_{(a_j, p_j) \in Y_i} P((a_j, p_j) | X_i) \quad (1)$$

The following section 2.2 and section 2.3 will introduce our MRC modeling for AE and prompt learning modeling for SC. An example of our modeling is given in Figure 2.

The extractive MRC task requires to extract the answer $A = \{a_1, a_2, \dots, a_m\}$ from a passage $G = \{g_1, g_2, \dots, g_n\}$ by answering a given query Q and each a_i in A is a span of G . The goal of a given model \mathcal{M} is to maximize the likelihood:

$$P(A | G, Q) = \prod_{a_i \in A} P(a_i | G, Q) \quad (2)$$

In this paper, we model AE as an extractive MRC task. For a sentence $X = \{x_1, x_2, \dots, x_n\}$ whose aspects are $A = \{a_1, a_2, \dots, a_m\}$ where a_i is a span of X , we regard X as the passage G and each sentiment aspect a_i as a corresponding answer. Following the procedure of MRC, we desire to find each sentiment aspect a_i by asking model \mathcal{M} a query Q . However, the query for the standard MRC task is given by the datasets, but AE datasets do not contain such an element. To let our task inputs compatible to the adopted MRC model, we construct a dedicated query Q . A large number of studies also have shown that the query has a significant impact on the performance of MRC no matter if such a query keeps a meaningful input or not. Following (Liu et al., 2021), we set a fixed-initialized query Q whose embedding can be optimized during the training process and search for the optimal best matching query in a continuous space for each sentence X .

2.2 Aspect Term Extraction as Machine Reading Comprehension

In detail, our MRC module takes a PrLM M as a backbone. The input X will be transformed into word embedding $W = \{w_1, w_2, \dots, w_n\}$ by the embedding layer \mathcal{E} of M . Then we initialize a query $Q = \{q_1, q_2, \dots, q_m\}$. Since the alternative query can largely influence the performance of the MRC, we add an encoder module to more effectively capture the optimal potential query embedding. The query encoder module consists of a randomly initialized embedding layer \mathcal{E}' , a Bi-direction LSTM layer and a double-layer MLP activated by RELU function (Glorot et al., 2011). The embedding r_i of q_i can be formalized as:

$$r_i = \text{MLP}([\text{BiLSTM}(\mathcal{E}'(q_{0:i})) : \text{BiLSTM}(\mathcal{E}'(q_{i:m}))]) \quad (3)$$

Then we combine the encoded embedding of query and sentence to form the input sequence for M . The embeddings of two special tokens [CLS] and [SEP] are bound to be inserted in the sequence. The final input sequence is like:

$$\{\mathcal{E}([\text{CLS}]), r_1, r_2, \dots, r_m, \mathcal{E}([\text{SEP}]), w_1, w_2, \dots, w_n\}$$

After feeding the sequence into M , the context representation $S \in \mathbb{R}^{n \times \|\mathcal{V}\|}$ of W is the only out-

put we need for the next steps, where the \mathcal{V} is the vocabulary of PrLM.

Subsequently, we carry out the selection of aspects. We adopt two independent binary classifiers to predict whether a token is a start or end position of an aspect following (Li et al., 2020). For the start position prediction, we first project the S into the dimension of $\mathbb{R}^{n \times 2}$ by a learnable weight $O_{start} \in \mathbb{R}^{\|\mathcal{V}\| \times 2}$ and get S' . Then, we apply the softmax to each row of S' to form a probability distribution which indicates the probability of every token to be the start word of an aspect. The above process can be formalized as:

$$P_{start} = \text{softmax}_{\text{each row}}(S \cdot O_{start}) \in \mathbb{R}^{n \times 2} \quad (4)$$

Following the same process, the probability distribution P_{end} of whether a token to be the end position of an aspect can be attained by the learnable weight $O_{end} \in \mathbb{R}^{\|\mathcal{V}\| \times 2}$.

The next step is to match the start position and end position in order to extract the final aspects. By the P_{start} and P_{end} , we can get the start as well as end positions simply according to the argmax function and store them into two sets, these are:

$$\begin{aligned} I_{start} &= \{i \mid \text{argmax}(P_{start}^{(i)}) = 1, i = 1, \dots, n\} \\ J_{end} &= \{j \mid \text{argmax}(P_{end}^{(j)}) = 1, j = 1, \dots, n\} \end{aligned} \quad (5)$$

The start position $i_{start} \in I_{start}$ means the $x_{i_{start}}$ should be the start token of an aspect and the $j_{end} \in J_{end}$ implies the end token $x_{j_{end}}$ of an aspect. Thus, we will train a sigmoid classifier to predict the match possibility of the i_{start}, j_{end} to be the boundary of one aspect, that is:

$$P_{i_{start}, j_{end}} = \text{sigmoid}(m \cdot \text{concat}(S_{i_{start}}, S_{j_{end}})) \quad (6)$$

where $m \in \mathbb{R}^{1 \times 2 \times \|\mathcal{V}\|}$.

During the training process, we leverage the CrossEntropy loss as below:

$$\begin{aligned} \mathcal{L}_{star} &= \text{CrossEntropy}(P_{start}, G_{start}) \\ \mathcal{L}_{end} &= \text{CrossEntropy}(P_{end}, G_{end}) \\ \mathcal{L}_{match} &= \text{CrossEntropy}(P_{start, end}, G_{start, end}) \end{aligned} \quad (7)$$

where the G_{start}, G_{end} and $G_{start, end}$ represent the golden labels. The total loss of this MRC module is the weighted sum of the above three losses:

$$\mathcal{L}_{MRC} = \alpha \mathcal{L}_{star} + \beta \mathcal{L}_{end} + \gamma \mathcal{L}_{match} \quad (8)$$

where the $\alpha, \beta, \gamma \in [0, 1]$ are three hyper-parameters to control the contributions of each objective function.

2.3 Aspect-level Sentiment Classification as Prompt Learning

Incorporating the output from the MRC module, we model the SC subtask as prompt learning, which allows us to transfer a classification problem into the form of predicting the [MASK] token contained in a prompt sentence Q as a pre-specified word w . Namely, for a given text X and its label Y , the purpose of prompt learning model \mathcal{M} is to maximize the likelihood:

$$P(Y | X) = P([\text{MASK}] = w | X, Q) \quad (9)$$

For a sentence X and one of its aspect a_j , we insert a_j into a pre-defined template $T = t_1, \dots, a_j, \dots, [\text{MASK}], \dots, t_m$ to form a prompt Q . Then we feed the sequence [CLS] X [SEP] Q into a module \mathcal{M} to measure the likelihood of a_j to be classified as polarity p_j :

$$P(p_j | a_j, X) = P([\text{MASK}] = w | [\text{CLS}]X[\text{SEP}]Q) \quad (10)$$

In detail, first of all, we need to construct the templates of prompts cautiously as the performance of prompt learning is very sensitive to the choice of prompts. Following (Seoh et al., 2021), we select four efficient prompt templates:

- $T_0 = \text{I felt the } a_i \text{ was } [\text{MASK}] .$
- $T_1 = \text{I } [\text{MASK}] \text{ the } a_i .$
- $T_2 = \text{The } a_i \text{ made me feel } [\text{MASK}] .$
- $T_3 = \text{The } a_i \text{ is } [\text{MASK}] .$

When given a sentence X and one of its aspect a_j , we fill $T_i (i \in \{1, 2, 3, 4\})$ to get four prompt Q_i and feed the sequence [CLS] X [SEP] Q_i into a PrLM, thus to attain its context representation C_i .

All we need is the representation $C_{i, [\text{MASK}]} \in \mathbb{R}^{\|\mathcal{V}\|}$ of [MASK] which contains the prediction information. To aggregate the prediction outcomes of the four prompts, we simply add them all up:

$$C_{[\text{MASK}]} = \sum_{i=0}^3 C_{i, [\text{MASK}]} \quad (11)$$

The current methods (Seoh et al., 2021; Zhang et al., 2021) directly take the probabilities of predicting [MASK] as several representative words to extract final sentiment polarities. For instance, Seoh et al. (2021) map the probabilities of three words $\{good, bad, ok\}$ to the probabilities of three sentiment polarities $\{positive, negative, neutral\}$ towards the template T_0 :

$$\begin{aligned} P([\text{MASK}] = good | X, a_i) &= P(p_i = positive | X, a_i) \\ P([\text{MASK}] = bad | X, a_i) &= P(p_i = negative | X, a_i) \\ P([\text{MASK}] = ok | X, a_i) &= P(p_i = neutral | X, a_i) \end{aligned}$$

This seems to be an extremely blunt approach, but we argue that it will increase the sparsity of the model and can not plausibly exert all the intelligence of $C_{[\text{MASK}]}$. Following such practice, we should exhaust all the tokens that can indicate sentiment to ameliorate this drawback. For example, if we want to predict the positive polarity, in addition to *good*, we also need to consider the words such as *nice, excellent, perfect*, etc. Obviously, such tedious work is not acceptable. In that case, we propose to utilize a double-layer MLP head activated by a RELU function to address the limitations mentioned above:

$$P(p_i | a_i, X) = \text{MLP}(C_{[\text{MASK}]}) \quad (12)$$

where the input dimension of MLP is $\|\mathcal{V}\|$ and the output dimension is set to 3 indicating the probability distribution over three polarities.

At this point, we merely need to follow the idea of the maximum likelihood method and apply CrossEntropy loss to calculate \mathcal{L}_{PL} for prompt learning, thus training MLP while fine-tuning BERT.

2.4 Training

The above MRC model with prompt learning can be trained together as a multi-task learning. We aggregate the two losses together for conducting back propagation and the total loss can be formalized as:

$$\mathcal{L}_{Total} = \mathcal{L}_{MRC} + \mathcal{L}_{PL} \quad (13)$$

2.5 Inference

When given trained MRC model for inferencing the AE result, the start and end positions are separately decided according to I_{start} and J_{end} (Eq.5) firstly. The following sigmoid classifier (Eq.6) will detect

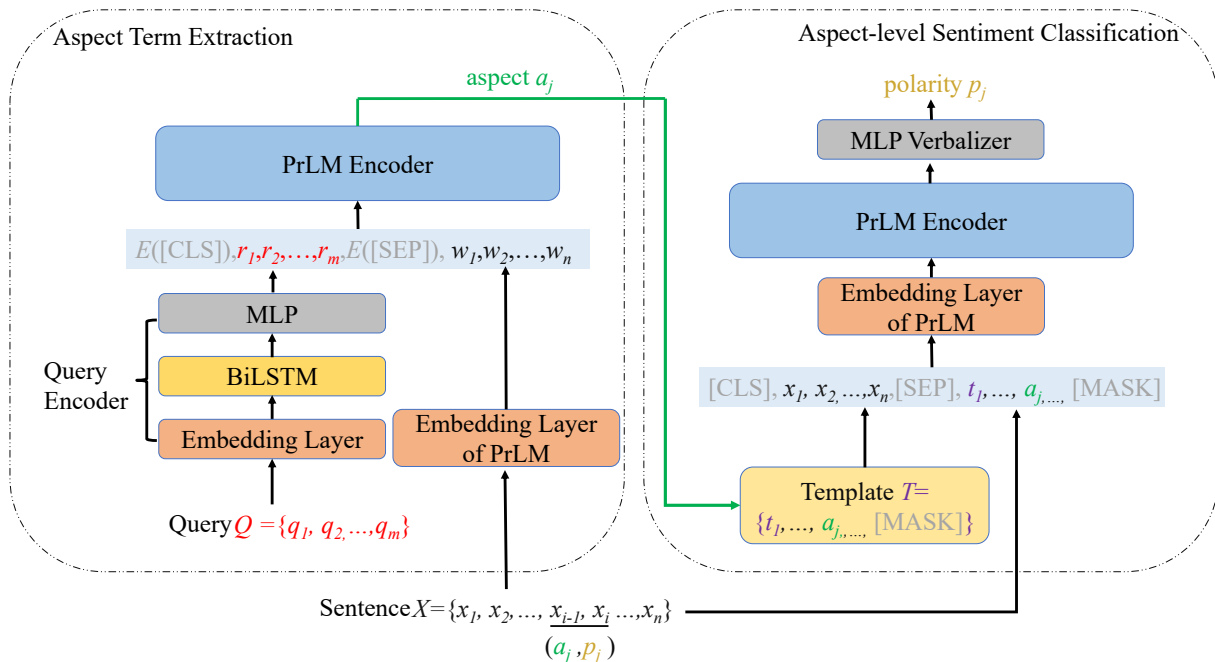


Figure 3: The architecture of our proposed MRCOOL framework.

	Lap14		Res14		Res15	
	#s	#a	#s	#a	#s	#a
train	3048	2373	3044	3699	1315	1199
test	800	654	800	1134	685	542

Table 1: The statistics of the three datasets (Wang et al., 2017). #s and #a denote the numbers of sentences and aspect terms.

the final start-end position combinations by I_{start} and J_{end} . As for inferring the SC result, the final polarity \hat{p} of an aspect is:

$$\begin{aligned} \hat{p} &= \operatorname{argmax} P(p_i | X, a_i) \\ &= \operatorname{argmax} \operatorname{MLP}(C_{[MASK]}) \end{aligned} \quad (14)$$

For AESC, the above two inference processes are united to obtain the final result. When given a sentence X , our framework first inputs it into the MRC model and receives the candidate aspect terms set A . Then, each a_i in A is enumerated to construct four templates with its homologous X . The prompt learning module takes them in and outputs the polarity \hat{p} , thus we get the triplet (*sentence, aspect, polarity*) which is served as the result of AESC.

3 Experiments

3.1 Setup

Datasets We conduct experiments on three widely used benchmark datasets derived from SemEval

2014 (Pontiki et al., 2014a) and SemEval 2015 (Pontiki et al., 2014b). For each benchmark, the golden boundaries of aspect terms are labeled and the aspect terms are annotated with positive, negative, or neutral polarities. So AE, SC and AESC subtasks are all available. LAPTOP2014 (Lap14) contains the reviews of the products from the laptop domain. RESTAURANT2014 (Res14) and RESTAURANT2015 (Res15) give some comments on foods and dining halls. The training/test splits are fixed for three datasets and more details about them are shown in Table 1.

Metrics For all experiments, we adopt F1 score as evaluation metric following the previous researches. For AE and SC, a predicted aspect term or polarity is correct only if it matches the golden data. And for AESC, we regard it as a right prediction only if an aspect term and its corresponding polarity are both recognized accurately at the same time.

PrLM and Settings For the fair comparison, our selected PrLMs are consistent with the previous strong baselines. We apply the publicly available BERT-Base-Uncased and BERT-Large-Uncased models² with the vanilla parameters and sizes for our MRCOOL framework. We adopt AdamW optimizer with the learning rate of $2e-5$ and warmup over the first 15% steps to train for 3 epochs. We use 30 epochs to train our framework.

²<https://github.com/google-research/bert>

	Lap14			Res14			Res15		
	AE	SC	AESC	AE	SC	AESC	AE	SC	AESC
IMN-BERT	77.35	75.56	61.73	84.06	75.67	70.72	69.90	70.10	60.22
SPAN-BERT	82.34	62.50	61.25	86.71	71.75	73.68	74.63	50.28	62.29
RACL-BERT	81.79	73.91	63.40	86.38	81.61	75.42	73.99	74.91	66.05
DUAL-MRC	82.51	75.97	65.94	86.60	82.04	75.95	75.08	73.59	65.08
BART-ABSA	82.52	76.76	67.37	87.07	75.56	73.56	75.48	73.91	66.61
MRCOOL	86.50	75.78	69.47	88.31	79.41	77.12	77.35	70.76	65.62

Table 2: Main results on three benchmark datasets for AE, SC and AESC. All results are measured by F_1 . The state-of-the-art results are in bold.

The batch size is 16 and the α, β, γ for MRC are all set to 1/3. For each experiment, we train our framework multiple times with different random seeds. The average of the best three results is regarded as a final result. We conduct all experiments on one Nvidia Titan RTX GPU.

3.2 Baselines

We compare our proposed MRCOOL framework with the following methods on AE, SC and AESC subtasks:

RACL-BERT Chen and Qian (2020) propose a RACL framework which stacks multiple layers. They also propose a relation propagation approach to obtain interactive signals among different subtasks. With the BERT model, their framework achieves good performance on AE, SC and AESC.

IMN-BERT He et al. (2019) put forward an end-to-end multi-task learning model for AE, SC and AESC. They apply a mechanism of information transmission to enhance their model.

SPAN-BERT Hu et al. (2019) propose a pipeline model for AESC. They apply BERT as their backbone network and a multi-target extractor is used to detect the boundaries of the sentiment aspects. Then a polarity classifier recognizes the polarity for each aspect.

DUAL-MRC Mao et al. (2021) present a unified framework for AESC. Two BERT models are contained by their framework and two different MRC tasks are carried out by them separately. The left BERT recognizes the boundaries of aspect terms and the right BERT extracts their polarities.

BART-ABSA Yan et al. (2021) propose the current state-of-art model which can solve the AE, SC and AESC. It redefines each subtask as a sequence mixed by pointer indexes and sentiment class in-

dexes. Then they convert all ABSA subtasks into a unified generative formulation. Finally, they use pre-trained sequence-to-sequence model BART to handle the subtasks in an end-to-end framework.

3.3 Main Results

Table 2 compares our results with other state-of-the-art approaches on three benchmark datasets. The results of AE are all obtained by MRC on BERT-large model. As for SC, we obtain the best results by taking the BERT-base model as the backbone of prompt learning for LAPTOP2014, RESTAURANT2015 datasets and the BERT-large model for RESTAURANT2014. More about the selection of different scale PrLM will be discussed in section 4.1. It is worth noting that our better results do not derive from the better PrLM, owing to DUAL-MRC (Mao et al., 2021) having already adopted the BERT-large model.

For the AE subtask, our MRC method has made good progress. We exceed the prior SOTA model by +3.98%, +1.24% and +1.87% on Lap14, Res14 and Res15 respectively. For the SC subtask, with our prompt learning method, we also achieve performance comparable to the best results before. For AESC subtask, our framework attains state-of-the-art performance by considerable margins over previous methods on Lap14 and Res14. Even if we do not obtain the SOTA on Res15, we also obtain a better result than most previous models.

The above results prove that our framework is very effective. We directly use the knowledge of the PrLM to avoid complex neural layers and feature engineering to attain SOTA or the results close to SOTA on multiple subtasks. The fine-tuning of PrLM is also very time-saving. With only 30 training epochs, our framework can get such a good

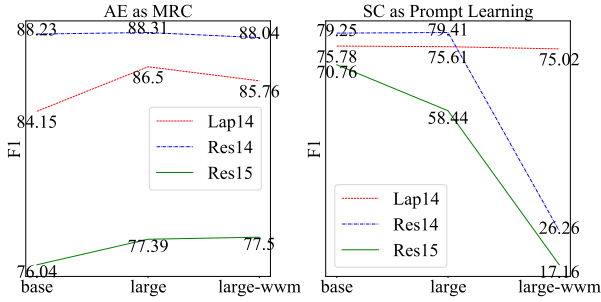


Figure 4: The F_1 change curve with the increment of the scale of PrLM on AE and SC.

	Lap14	Res14	Res15
w/o Query Encoder	85.24	87.35	77.05
Query Encoder	86.50	88.31	77.35

Table 3: Ablation study on Query Encoder for MRC. The BERT-large model is adopted. All results are F_1 for the AE subtask.

performance. The experimental results indicate the effectiveness and simplicity of our MRCOOL framework.

4 Ablation Study

4.1 Effect of different PrLM scale

Plenty of earlier studies have proved that the scale of the PrLM has a great influence on the performance of downstream tasks. Therefore, we adopt BERT-base, BERT-large and BERT-large-wwm models that derived from one series but of different scales to test our framework. The results are given in Figure 4.

For MRC, it can be speculated from the curve that the PrLM of different scales is relatively stable on the three datasets. Among three PrLM, the performance of BERT-large is usually better than BERT-base and BERT-large-wwm. For prompt learning, some not robust phenomena happen. BERT-base still performs well on the three datasets. But with the BERT-large, a significant loss of accuracy on the Res15 occurs. And BERT-large-wwm produces disastrous results both on Res14 and Res15. We check the training logs of these poorly experiments and find that the training loss usually reaches zero after 10 epochs but the testing loss still maintains a high value. This means that heavy overfitting could be triggered.

For these three benchmark datasets, BERT-large seems to be usually the best choice and it is definitely not true that a larger scale of PrLM leads to

	Lap14	Res14	Res15
w/o Prompt Encoder	75.78	79.25	70.76
Prompt Encoder	73.92	73.04	66.55

Table 4: Ablation study on Prompt Encoder for prompt learning. The BERT-large model is adopted. All results are F_1 for the SC subtask.

	Lap14	Res14	Res15
Selecting Words	74.40	78.95	69.45
Single-Layer MLP	74.69	79.21	69.29
Double-Layer MLP	75.78	79.25	70.76

Table 5: Ablation study on different verbalizer. Selecting words means the current dosage of selecting the probability distribution of a few of words. The BERT-large model is applied and all results are F_1 for the AE subtask.

a better result.

4.2 Effect of Query/Prompt Encoder Module

We propose a query encoder module in this paper to find the potential optimal query for MRC. We also directly think of whether we can learn a potential optimal prompt template by the prompt encoder module whose structure is the same as the query encoder module. For example, we desire to find the optimal embedding of *I felt the a_i was [MASK]* in the template T_0 . So we conduct ablation studies on the query encoder of MRC module and prompt encoder of the prompt learning module, respectively. The results are shown in the Table 3 and Table 4.

It can be indicated that the introduced query encoder has facilitated the performance of MRC, which is in line with the previous research that the query plays an important role in MRC and also supports the effectiveness of the query encoder module. But the prompt encoder harms to the prompt learning. The explanation we give is that we use four different templates for prompt learning and it is too difficult for a simple query encoder module to learn perfect embeddings for them all. We attempted to apply only one template and combine the query encoder module. The result shows that the query encoder module does boost the performance but such a manner is far not as good as the combination of four templates. Thus, we abandon the query encoder module in our framework.

4.3 Effect of MLP Verbalizer

In this paper, one of our major improvements to prompt learning is to desert the original verbalizer that selecting the probability distribution of minority specific words from the output word-embedding of [MASK] and we leverage an MLP with a RELU activation function to overcome the sparsity and loss of information. We respectively use the original method, a single-layer MLP and a double-layer MLP with RELU activation function for ablation study. Table 5 displays the result.

Experiments express that MLP does have a relatively large improvement in the effect of prompt learning and the double-layer MLP with RELU has a stronger ability to learn the probability distribution of three polarities. As a result, our proposed simple MLP verbalizer shows quite effective.

5 Related Work

Open-domain sentiment analysis or ABSA requires to extracting the aspect terms with their corresponding sentiment polarities in the open domain, which is an active research topic in recent years. Early studies treat them as two separate tasks and use some traditional algorithms such as Conditional Random Fields (CRF) to complete the task. Wang et al. (2016) apply recursive neural CRF to perform ABSA. Shu et al. (2017) use a lifelong learning CRF to extract the prior knowledge of past domains. With the rise of deep learning technology, more and more models based on neural networks have begun to emerge. (Poria et al., 2016; Xu et al., 2018; Shu et al., 2019; Wu et al., 2021) take convolutional neural network (CNN) to handle ABSA tasks. In addition, some researchers apply RNN and also make some progress (Wadawadagi and Pagi, 2018; Han et al., 2018; Luo et al., 2019; Zeng et al., 2019). In recent years, more studies have proposed diverse attention mechanisms to extract more knowledge from the text and boost the performance of the model. Wang et al. (2017) put forward a novel layer containing two attentions to extract sentiment aspects, opinions and polarities. Li et al. (2018) propose a history attention to exploit the opinion summary and the aspect detection history. Rida-E-Fatima et al. (2019) propose a deep learning-based multilayer dual-attention model to extract the mediate relationships between the aspects and opinions.

Recently, offering promising performance, PrLM has become an important and rapid devel-

opment area in the field of natural language processing. However, minority studies directly utilize the knowledge of PrLM. Even if the current state-of-art model (Yan et al., 2021), it merely regards BART (Lewis et al., 2020) as a powerful seq-to-seq model without using its erudition. For extracting the knowledge from PrLM, MRC is an excellent solution. Li et al. (2020) convert named entity recognition into an MRC task and achieve the state-of-the-art. Mao et al. (2021) design a DUAL-MRC framework for ABSA and get the promising result. Gan et al. (2021) employ MRC to handle dependency parsing. Yu et al. (2021) and Chen et al. (2021a) respectively propose a self question-answering model and a bidirectional MRC model for ABSA, but they can not solve all three AE, SC and AESC tasks.

For prompt learning, many researchers regard it as a new learning paradigm along with the swift growth of PrLM and argue that it can effectively reel off the enlightenment of PrLM. (Chen et al., 2021b; Seoh et al., 2021) have applied prompt learning to named entity recognition and sentiment analysis respectively. Han et al. (2021) implement a sentence classification model by prompt learning and logic rules. On the selection and generation of prompt, Shin et al. (2020) propose an automatic prompt generation method and Liu et al. (2021) put forward a p-tuning idea to improve the effectiveness of prompt.

However, the above works do not pay attention to the query generation of MRC and the sparsity of the verbalizer of prompt learning. In this paper, we propose a MRCOOL framework for ABSA which designs a query encoder to improve the capability of MRC and a simple MLP verbalizer is used to reduce the sparsity of prompt learning.

6 Conclusion

In this paper, we propose a MRCOOL framework to handle AE, SC and AESC subtasks of ABSA in one shot through the process of MRC with prompt learning. In detail, we first model aspect extraction as an MRC task and then let the MRC module help aspect-level sentiment classification implemented in a prompt learning way so that we present an end-to-end framework to fulfill the complete task requirement of ABSA. The experimental results demonstrate the effectiveness of our proposed framework by providing consistent and general performance improvement over strong baselines. In

detail, our framework attains new state-of-the-art for AE subtask by considerable margins over previous methods on three datasets and reaches state-of-the-art performance for AESC on two datasets.

References

- Shaowei Chen, Jie Liu, Yu Wang, Wenzheng Zhang, and Ziming Chi. 2020. [Synchronous double-channel recurrent network for aspect-opinion pair extraction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6515–6524, Online. Association for Computational Linguistics.
- Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021a. [Bidirectional machine reading comprehension for aspect sentiment triplet extraction](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 12666–12674.
- Xiang Chen, Ningyu Zhang, Lei Li, Xin Xie, Shumin Deng, Chuanqi Tan, Fei Huang, Luo Si, and Hua-jun Chen. 2021b. [Lightner: A lightweight generative framework with prompt-guided attention for low-resource ner](#). *ArXiv preprint*, abs/2109.00720.
- Xiang Chen, Ningyu Zhang, Xin Xie, Shumin Deng, Yunzhi Yao, Chuanqi Tan, Fei Huang, Luo Si, and Huajun Chen. 2021. [KnowPrompt: Knowledge-aware Prompt-tuning with Synergistic Optimization for Relation Extraction](#). *arXiv e-prints*, page arXiv:2104.07650.
- Zhuang Chen and Tiejun Qian. 2020. [Relation-aware collaborative learning for unified aspect-based sentiment analysis](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3685–3694, Online. 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. Association for Computational Linguistics.
- Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. [Multi-grained attention network for aspect-level sentiment classification](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442, Brussels, Belgium. Association for Computational Linguistics.
- Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. [Target-oriented opinion words extraction with target-fused neural sequence labeling](#). 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 2509–2518, Minneapolis, Minnesota. Association for Computational Linguistics.
- Leilei Gan, Yuxian Meng, Kun Kuang, Xiaofei Sun, Chun Fan, Fei Wu, and Jiwei Li. 2021. [Dependency parsing as mrc-based span-span prediction](#). *ArXiv preprint*, abs/2105.07654.
- Michael Glass, Alfio Gliozzo, Rishav Chakravarti, Anthony Ferritto, Lin Pan, GP Bhargav, Dinesh Garg, and Avirup Sil. 2019. [Span selection pre-training for question answering](#). *arXiv preprint arXiv:1909.04120*.
- Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. [Deep sparse rectifier neural networks](#). In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings.
- Xu Han, Weilin Zhao, Ning Ding, Zhiyuan Liu, and Maosong Sun. 2021. [Ptr: Prompt tuning with rules for text classification](#). *ArXiv preprint*, abs/2105.11259.
- Zhongming Han, Xin Jiang, Mengqi Li, Mengmei Zhang, and Dagao Duan. 2018. [An integrated semantic-syntactic sbLSTM model for aspect specific opinion extraction](#). In *International Conference on Web Information Systems and Applications*, pages 191–199. Springer.
- Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. [An interactive multi-task learning network for end-to-end aspect-based sentiment analysis](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 504–515, Florence, Italy. Association for Computational Linguistics.
- Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. [Open-domain targeted sentiment analysis via span-based extraction and classification](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.
- Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. [BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.
- Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020. [A unified MRC framework for named entity recognition](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5849–5859, Online. Association for Computational Linguistics.

- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. [A unified model for opinion target extraction and target sentiment prediction](#). In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 6714–6721. AAAI Press.
- Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018. [Aspect term extraction with history attention and selective transformation](#). In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 4194–4200. ijcai.org.
- Xin Li and Wai Lam. 2017. [Deep multi-task learning for aspect term extraction with memory interaction](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2886–2892, Copenhagen, Denmark. Association for Computational Linguistics.
- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. [Event extraction as machine reading comprehension](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651, Online. Association for Computational Linguistics.
- Xiao Liu, Yanan Zheng, Zhengxiao Du, Ming Ding, Yujie Qian, Zhilin Yang, and Jie Tang. 2021. [Gpt understands, too](#). *ArXiv preprint*, abs/2103.10385.
- Yang Liu, Sujian Li, Xiaodong Zhang, and Zhifang Sui. 2016. [Implicit discourse relation classification via multi-task neural networks](#). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 2750–2756. AAAI Press.
- Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019. [DOER: Dual cross-shared RNN for aspect term-polarity co-extraction](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 591–601, Florence, Italy. Association for Computational Linguistics.
- Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. [A joint training dual-mrc framework for aspect based sentiment analysis](#). *ArXiv preprint*, abs/2101.00816.
- Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2020. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8600–8607.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014a. [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. Association for Computational Linguistics.
- Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014b. [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. Association for Computational Linguistics.
- Soujanya Poria, Erik Cambria, and Alexander Gelbukh. 2016. Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, 108:42–49.
- Syeda Rida-E-Fatima, Ali Javed, Ameen Banjar, Aun Irtaza, Hassan Dawood, Hussain Dawood, and Abdullah Alamri. 2019. [A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis](#). *IEEE Access*, 7:114795–114807.
- Ronald Seoh, Ian Birlle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. 2021. [Open aspect target sentiment classification with natural language prompts](#). *ArXiv preprint*, abs/2109.03685.
- Taylor Shin, Yasaman Razeghi, Robert L. Logan IV, Eric Wallace, and Sameer Singh. 2020. [AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 4222–4235, Online. Association for Computational Linguistics.
- Lei Shu, Hu Xu, and Bing Liu. 2017. [Lifelong learning CRF for supervised aspect extraction](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 148–154, Vancouver, Canada. Association for Computational Linguistics.
- Lei Shu, Hu Xu, and Bing Liu. 2019. [Controlled cnn-based sequence labeling for aspect extraction](#). *ArXiv preprint*, abs/1905.06407.
- Chuangdong Su, Fumiyo Fukumoto, Xiaoxi Huang, Jiyi Li, Rongbo Wang, and Zhiqun Chen. 2020. [DeepMet: A reading comprehension paradigm for token-level metaphor detection](#). In *Proceedings of the Second Workshop on Figurative Language Processing*, pages 30–39, Online. Association for Computational Linguistics.
- Yi Sun, Yu Zheng, Chao Hao, and Hangping Qiu. 2021. [NSP-BERT: A Prompt-based Zero-Shot Learner Through an Original Pre-training Task-Next Sentence Prediction](#). *arXiv e-prints*, page arXiv:2109.03564.

- Ramesh S Wadawadagi and Veerappa B Pagi. 2018. A deep recursive neural network model for fine-grained opinion classification. In *International Conference on Recent Trends in Image Processing and Pattern Recognition*, pages 607–621. Springer.
- Hai Wan, Yufei Yang, Jianfeng Du, Yanan Liu, Kunxun Qi, and Jeff Z Pan. 2020. Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 9122–9129.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016. [Recursive neural conditional random fields for aspect-based sentiment analysis](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 616–626, Austin, Texas. Association for Computational Linguistics.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2017. [Coupled multi-layer attentions for co-extraction of aspect and opinion terms](#). In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, pages 3316–3322. AAAI Press.
- Chao Wu, Qingyu Xiong, Zhengyi Yang, Min Gao, Qiude Li, Yang Yu, Kaige Wang, and Qiwu Zhu. 2021. Residual attention and other aspects module for aspect-based sentiment analysis. *Neurocomputing*, 435:42–52.
- Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, and Jiwei Li. 2019. Coreference resolution as query-based span prediction. *arXiv preprint arXiv:1911.01746*.
- Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. [Double embeddings and CNN-based sequence labeling for aspect extraction](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 592–598, Melbourne, Australia. Association for Computational Linguistics.
- Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. [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, Online. Association for Computational Linguistics.
- Guoxin Yu, Jiwei Li, Ling Luo, Yuxian Meng, Xiang Ao, and Qing He. 2021. Self question-answering: Aspect-based sentiment analysis by role flipped machine reading comprehension. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1331–1342.
- Jianfei Yu, Jing Jiang, and Rui Xia. 2018. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 27(1):168–177.
- Jiangfeng Zeng, Xiao Ma, and Ke Zhou. 2019. Enhancing attention-based LSTM with position context for aspect-level sentiment classification. *IEEE Access*, 7:20462–20471.
- Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. [Aspect sentiment quad prediction as paraphrase generation](#). *ArXiv preprint*, abs/2110.00796.
- Zhuosheng Zhang, Junjie Yang, and Hai Zhao. 2020. Retrospective reader for machine reading comprehension. *arXiv preprint arXiv:2001.09694*.