Self Question-answering: Aspect Sentiment Triplet Extraction via a Multi-MRC Framework based on Rethink Mechanism

Fuyao Zhang[1],Yijia Zhang[1,✉],Mengyi Wang[1],Hong Yang[1],Mingyu Lu[1],Liang Yang[2]
[1]Dalian Maritime University, Dalian
{zhangfuyao,zhangyijia,mengyiw,yanghong,lumingyu}@dlmu.edu.cn
[2]Dalian University of Technology, Dalian
liang@dlut.edu.cn

Abstract

The purpose of Aspect Sentiment Triplet Extraction (ASTE) is to extract a triplet, including the target or aspect, its associated sentiment, and related opinion terms that explain the underlying cause of the sentiment. Some recent studies fail to capture the strong interdependence between ATE and OTE, while others fail to effectively introduce the relationship between aspects and opinions into sentiment classification tasks. To solve these problems, we construct a multi-round machine reading comprehension framework based on a rethink mechanism to solve ASTE tasks efficiently. The rethink mechanism allows the framework to model complex relationships between entities, and exclusive classifiers and probability generation algorithms can reduce query conflicts and unilateral drops in probability. Besides, the multi-round structure can fuse explicit semantic information flow between aspect, opinion and sentiment. Extensive experiments show that the proposed model achieves the most advanced effect and can be effectively applied to ASTE tasks.

1 Introduction

Aspect-based Sentiment Analysis (ABSA) is a fine-grained task (Zhang et al., 2022). Its purpose is to detect the sentiments of different entities rather than infer the overall sentiment of sentences. As shown in Figure 1, researchers proposed many subtasks of ABSA, such as Aspect Term Extraction (ATE) (Ma et al., 2019), Opinion Term Extraction (OTE) (Zhao et al., 2020), Aspect-based Sentiment Classification (ABSC) (Hazarika et al., 2018), Aspect-oriented Opinion Extraction (AOE) (Fan et al., 2019), etc. Aspect terms refer to words or phrases that describe the attributes or characteristics of an entity. Opinion terms refer to words or phrases that express the corresponding attitudes of the aspect terms. ATE and OTE aim to extract aspects and opinions from sentences, respectively. For ABSC, given a sentence and an aspect within the sentence, it is possible to predict the sentiment (positive, neutral, or negative) associated with that aspect. In the sentence "The service is good, but the food is not so great", ATE extracts "service" and "food", and OTE extracts "good" and "not so great". ABSC predicts the sentiment polarity of "service" and "food" as positive and negative, respectively. However, these studies focus on individual tasks respectively while neglecting their interdependencies.

Recent studies have focused on joint tasks to explore the interactions among different tasks. Figure 1 provides examples of Aspect Term Extraction and Sentiment Co-classification (AESC) as well as Aspect-Opinion Pair Extraction(pair). However, these subtasks still cannot tell a complete story. Hence Aspect Sentiment Triplet Extraction (ASTE) was introduced. The purpose of ASTE is to extract aspect terms, related opinion terms, and sentiment polarities for each aspect simultaneously. ASTE has two advantages: first, opinions can enhance the expressiveness of the model, helping to determine the sentiment of the aspects better; second, the sentiment dependency between aspects and opinions can narrow the gap of sentiment decision-making, further improving the interpretability of the model.

Peng (Peng et al., 2020) proposed the first solution for ASTE, which jointly extracts aspect-sentiment pairs and opinions using two sequence taggers. Sentiment is attached to aspects through a unified tagging

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

S: The service is good, but the food is not so great.

Substask			
Aspect Term Extraction(ATE)	S	⇨	{service, food}
Opinion Term Extraction(OTE)	S	⇨	{good, not so great}
Aspect-based Sentiment Classification(ABSC)	S+service	⇨	Positive
	S+food	⇨	Negative
Aspect-oriented Opinion Extraction(AOE)	S+service	⇨	good
	S+food	⇨	not so great
Aspect Term and Sentiment Co-extraction(AESC)	S	⇨	{service, Positive}
	S	⇨	{food, Negative}
Aspect-Opinion Pair Extraction(Pair)	S	⇨	{service, good}
	S	⇨	{food, not so great}
Aspect Sentiment Triplet Extraction(ASTE)	S	⇨	{service, good, Positive}
	S	⇨	{food, not so great, Negative}

Figure 1: Illustration of ABSA subtasks

process, and then an exclusive classifier is used to pair the extracted aspect-sentiment pairs with opinions. While this method achieved significant results, there are also some issues. **Firstly**, the model has low computational efficiency because its framework involves two stages and requires training three independent models. **Secondly**, the model does not fully recognize the relationship between ATE and OTE, and does not effectively utilize the correspondence between aspect terms and opinion terms. **Thirdly**, the correspondence between aspect and opinion expressions can be very complex, involving various relationships such as one-to-many, many-to-one, overlapping, and nesting, which makes it difficult for the model to flexibly and accurately detect these relationships. Therefore, we take the solution to the above problems as our challenge.

To address the **first** problem mentioned above, inspired by (Chen et al., 2021), this paper proposes an improved multi-round MRC framework (R-MMRC) with a rethink mechanism to elegantly identify ASTE within a unified framework. To address the **second** problem, we decompose the ASTE into multiple rounds and introduce prior knowledge from the previous round to the current round, which effectively learns the associations between different subtasks. In the first round, we design static queries to extract the first entity of each aspect-opinion pair. In the second round, we design dynamic queries to identify the second entity of each aspect-opinion pair based on the previously extracted entity. In the third round, we design a dynamic sentiment query to predict the corresponding sentiment polarity based on the aspect-opinion pairs obtained in the previous round. In each step, the manually designed static and dynamic queries fully utilize the sentence's explicit semantic information to improve the extraction or classification performance. Based on these steps, we can flexibly capture complex relationships between entities, effectively mine the connection between ATE and OTE, and use these relationships to guide sentiment classification. To address the **third** issue, inspired by human two-stage reading behaviour (Zheng et al., 2019), we introduce a rethink mechanism to validate candidate aspect-opinion pairs further, enhance the information flow between aspects and opinions, and improve overall performance. Our contributions are summarized as follows:

- We proposed an improved multi-round machine reading comprehension framework (R-MMRC) with a rethink mechanism to address the ASTE task effectively.
- The model introduced the rethink mechanism to enhance the information flow between aspects and opinions. The exclusive classifier was added to avoid interference and query conflicts between different Q&A steps. The probability generation algorithm was also introduced to improve the prediction performance further.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

- The proposed model conducts extensive experiments on four public datasets, and experimental results show that our framework is very competitive.

2 Related work

We present related work in two parts, including various subtasks of aspect-based sentiment analysis and machine reading comprehension.

2.1 Aspect-based Sentiment Analysis

ATE. Locating and extracting terms that are pertinent for sentiment analysis and opinion mining is the task of ATE (Xu et al., 2018). Recent studies use two ways to alleviate the noise from pseudo-labels generated by self-learning (Wang et al., 2021).

OTE. OTE is to extract opinion terms corresponding to aspect terms, hoping to find specific words or phrases that describe sentiment (Chen and Qian, 2020).

ABSC. The task's aim is to forecast sentiment polarity of specific aspects. The latest development of ABSC focuses on developing various types of deep learning models: CNN-based (Huang and Carley, 2019), memory-based methods (Majumder et al., 2018), etc. Dependencies and graph structures have also been used effectively for sentiment classification problems (Xu et al., 2020a; Zhang and Qian, 2020).

AOE. Fan (Fan et al., 2019) first proposed this subtask, which aims to extract corresponding opinion terms for each provided aspect term. The difference between AOE and OTE is that the input of AOE contains aspect terms.

AESC. AESC aims to simultaneously extract aspect terms and sentiment. Recent work removes the boundaries of these two subtasks using a unified approach. Chen (Chen and Qian, 2020) proposes a relational awareness framework that allows subtasks to coordinate their work by stacking multitask learning and association propagation mechanisms.

Pair. The Pair task usually uses the pipeline method or directly uses the unified model. Gao (Gao et al., 2021) proposed a machine reading comprehension task based on question answering and span annotation.

ASTE. Peng (Peng et al., 2020) defined a triplet extraction task intending to extract all possible aspect terms, their corresponding opinion terms, and sentiment polarities. Xu (Xu et al., 2021) propose a span-based method to learn the interaction between target words and opinion words and propose a two-channel span pruning strategy.

2.2 Solving NLP Tasks by MRC

The purpose of machine reading comprehension (MRC) is to enable machines to answer questions from a specific context based on queries. Xu (Xu et al., 2021) proposed a post-training method for BERT. Yu (Yu et al., 2021) introduced role replacement into the reading comprehension model and solved the coupling problem in different aspects. To sum up, MRC is an effective and flexible framework for natural language processing tasks.

2.3 Aspect Sentiment Triplet Extraction

ASTE is the latest subtask in the field of ABSA. Xu (Xu et al., 2020b) proposed a position-aware tagging scheme that efficiently captures interactions in triplets. However, they generally overlooked the relationship between words and language features. In a similar vein, Yan (Yan et al., 2021) converted the ASTE task into a generative formulation, but also tended to ignore the linguistic aspects of word features. Meanwhile, Chen (Chen et al., 2022) introduced an enhanced multi-channel GCN that incorporated various language features to enhance the model. However, they failed to consider the interaction between these different language features. In summary, there are still many issues waiting to be resolved in ASTE, and we will try our best to make breakthroughs in ASTE tasks.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

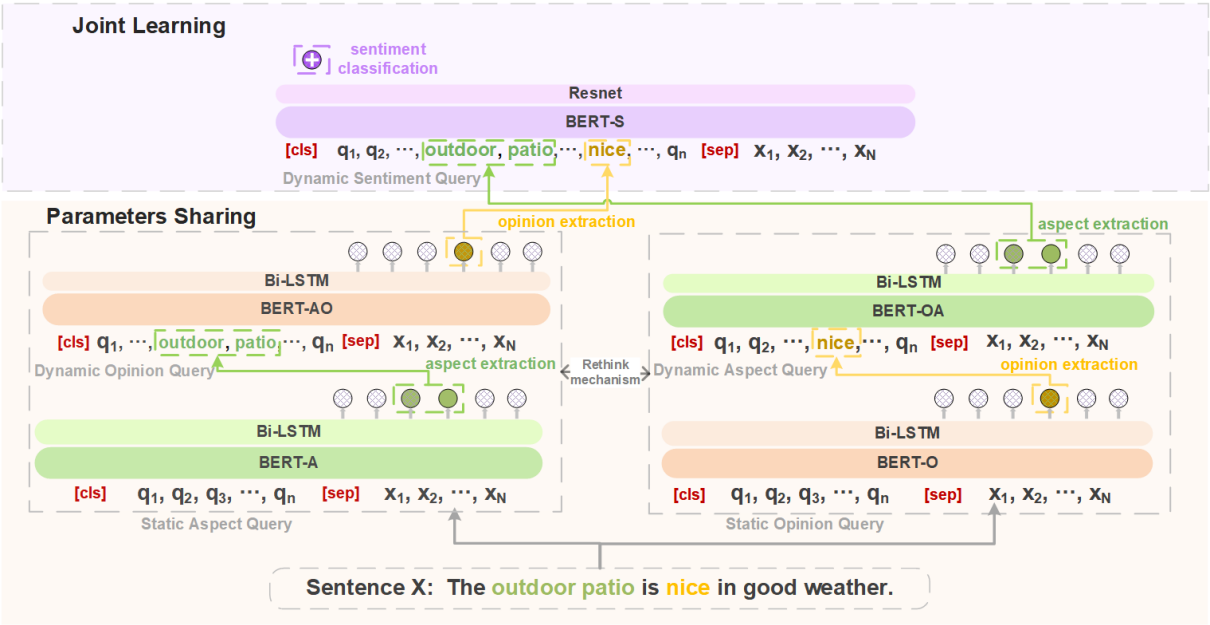

Figure 2: Overview of R-MMRC framework

3 Methodology

3.1 Model Framework

As shown in Figure 2, to address the ASTE task, we propose a multi-round machine reading comprehension framework based on a rethink mechanism. Specifically, we design two modules: parameter sharing and joint learning. First, for the parameter sharing module, we design a bidirectional structure to extract aspect-opinion pairs, consisting of two querying rounds. The first round is static queries aimed at extracting all aspect or opinion sets based on the given query statements. The second round is dynamic queries, aimed at identifying the corresponding opinion or aspect sets based on the results of the static queries and generating aspect-opinion pairs. Then, the rethink mechanism is used to filter out invalid aspect-opinion pairs in the parameter sharing stage. For the joint learning module, the framework employs dynamic sentiment queries to predict the sentiment polarity of the filtered aspect-opinion pairs. During the probability generation stage, the model combines the answers from different queries and forms triplets.

3.2 Query Template Construction

In R-MMRC, we build queries using a template-based method. Specifically, we designed static queries $Q^S = \{q_i^S\}_{i=1}^{|Q^S|}$ and dynamic queries $Q^D = \{q_i^D\}_{i=1}^{|Q^D|}$, where i represents the i-th token in the sentence. In particular, static queries do not carry any contextual information. Dynamic queries require the results of static queries as keywords to search for valid information in sentences. Static and dynamic queries are used to formalize the ASTE task as an MRC task:

Parameter Sharing.

Static Aspect Query q_A^S: We design the query ' Find the aspect in the text? ' to extract a set of aspects $A = \{a_i\}_{i=1}^{|A|}$ from a given review sentence X.

Dynamic Opinion Query q_O^D: We design the query ' Find the opinion of the aspect a_i ?' to extract the relevant opinions $O_{ai} = \{o_{ai,j}\}_{j=1}^{|O_{ai}|}$ for each aspect a_i.

Static Opinion Query q_O^S: We design the query ' Find the opinion in the text? ' to extract the collection of opinions $O = \{o_i\}_{i=1}^{|O|}$ from a given review sentence X.

Dynamic Aspect Query q_A^D: We design the query ' Find the aspect of the opinion o_i ' to extract the corresponding aspects $A_{oi} = \{a_{oi,j}\}_{j=1}^{|A_{oi}|}$ for each opinion O_i.

Through the above queries, dynamic queries elegantly learn the conclusions of static queries and

naturally integrates entity extraction and relationship detection. Although the entity results of these two queries are the same, the latter conveys the information of the former and searches for all entities described by the former, while the former does not carry any contextual information. Then, in the joint learning module, we classify the sentiment corresponding to the aspect-opinion pairs.

Joint Learning.

Dynamic Sentiment Query $q^{D'}$: We build the query ' Find the sentiment of the aspect a_i and the opinion o_i? ' to anticipate the sentiment polarity s_i of each aspect a_i.

Through the queries, we can fully consider the semantic relationship of aspect terms and corresponding opinion terms.

3.3 Input Representations

This section focuses on the triplet extraction task. Given a sentence $X = \{x_1, x_2, \ldots, x_N\}$ with max-length N as the input, and each query $q_i = \left\{q_1^i, q_2^i, \ldots, q_{|q_i|}^i\right\}$ with $|q_i|$ tokens. We use BERT as the model's encoder, and the encoding layer's role is to learn each token's context representation. First, we associate the query Q_i with the review sentence X and obtain the input $I = \left\{[CLS], q_1^i, q_2^i, \ldots, q_{|q_i|}^i, [SEP], x_1, x_2, \ldots, x_N\right\}$ after combination, where $[CLS]$ and $[SEP]$ are the start tag and the segment tag. Bert is used to encode an initial representation sequence $E = \left\{e_1, e_2, \ldots, e_{|q_i|+2+N}\right\}$,which is encoded as a hidden representation sequence $H_e = \left\{h_1, h_2, \ldots, h_{|q_i|+2+N}\right\}$ with stacked transformer blocks.

3.4 Query Answer Prediction

For the first two rounds of static and dynamic queries, the answer is to extract aspect terms or opinion terms from review sentence X. For instance, in Figure 2, the aspect term "outdoor patio" should be extracted as the answer to the Static Aspect Query.

In the original BMRC (Chen et al., 2021), all queries shared a single classifier, which could lead to interference between different types of queries and cause query conflicts. Since there are four different queries in the parameter sharing part, we set an exclusive BERT classifier for each query, which can effectively avoid interference of query conflict and answering step. Classifiers are BERT-A, BERT-AO, BERT-O, and BERT-OA, respectively. The context representation generated by BERT is used for Bi-LSTM to generate sentence hidden state vectors. Since H_e already contains information about aspect or opinion, we obtain specific context representation by aggregating the hidden states of two directions: $H = \left[\overrightarrow{H_{e_f}}; \overleftarrow{H_{e_b}}\right]$, where $\overrightarrow{H_{e_f}}$ is the hidden state of the forward LSTM and $\overleftarrow{H_{e_b}}$ is of the backward LSTM. We adopted the strategy of (Xu et al., 2019) and employ two binary classifiers to predict the answer spans based on the hidden representation sequence H. We utilize two classifiers to predict the possibility that the token x_i is the start or end of the answer. Then, we obtain the logits and probabilities for start and end positions:

$$p_{x_i,q}^{\text{start}} = \text{softmax}\left(W_s h_{|q|+2+i}\right) \tag{1}$$

$$p_{x_i,q}^{end} = \text{softmax}\left(W_e h_{|q|+2+i}\right) \tag{2}$$

where $W_s \in R^{d\times 2}$ and $W_e \in R^{d\times 2}$ are model parameters, d represents the dimension of hidden representations, and $|q|$ stands for the query length.

For dynamic sentiment queries, we utilize the hidden representation of [CLS] to predict the answer. We add a three-class classifier in BERT, called "BERT-S" for short, to predict the sentiment of aspect-opinion pairs. In addition, we add two layers of ResNet network to protect the integrity of information and reduce the loss of information.

$$h = \sigma F\left(h_1, \{W_{ri}\}\right) + h_1 \tag{3}$$

$$p_{X,q}^{D'} = \text{softmax}\left(W_c h\right) \tag{4}$$

where h_1 is the hidden representation of $[CLS]$, refers to ReLU activation function, $F()$ is the residual mapping of fitting, W_{ri} and $W_c = R^{d\times 3}$ is the model parameter.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

3.5 Rethink Mechanism

During the inference process, we combine the answers from different queries into tuples. As shown in Figure 2, the left-side static aspect query q_A^S first identifies all aspect items $A = \{a_1, a_2, \ldots, a_{|A|}\}$. For each aspect item a_i, the corresponding opinion expression set $O_i = \{o_{i,1}, o_{i,2}, \ldots, o_{i,|O_i|}\}$ is identified through the dynamic opinion query q_O^S, resulting in a set of aspect-opinion pairs $V_{AO} = \left[\left(a_i^k, o_{i,j}^k\right)\right]_{k=1}^{I}$, and ultimately obtaining the probability of each candidate pair $p\left(a_i, o_{i,j}\right) = p\left(a_i\right) p\left(o_{i,j} \mid a_i\right)$. Similarly, on the right side, the model first identifies all the opinion items and then queries all corresponding aspect items, and we finally obtain another set of aspect-opinion pairs $V_{OA} = \left[\left(a_{j,i}^k, o_j^k\right)\right]_{k=1}^{J}$, from which we obtain the probability of each candidate pair $p\left(a_{j,i}, o_j\right) = p\left(o_j\right) p\left(a_{j,i} \mid o_j\right)$.

However, the above approach may introduce incorrect aspect-opinion pairs. To better address this issue, we implement a rethink mechanism through a soft-selection strategy. If there exist identical candidate pairs in sets V_{AO} and V_{OA}, then the corresponding aspect-opinion pairs are added to the valid set V. If there are unmatched candidate pairs in V_{AO} and V_{OA}, it indicates that one side's output may be invalid. Therefore, in the soft selection strategy, we adjust the probabilities and introduce a probability threshold λ. If the probability $p(a, o)$ of a certain candidate pair in the difference set is greater than or equal to the probability threshold λ, then this candidate pair is added to the valid set V; otherwise, it is discarded. By using a rethink mechanism, invalid pairs can be better filtered out, reducing the interference of erroneous candidate pairs on the model.

3.6 Entity Pair Probability Generation

After filtering with the rethink mechanism, we obtained a set of valid aspect-opinion pairs, and the next step is to calculate the probability of each candidate pair. In BMRC, the probability of an entity is the product of the probabilities of its start and end positions, and the probability of a candidate pair is the product of the probabilities of the aspect item and opinion item. However, this can result in a product of high probabilities equaling a lower probability value, which does not well represent the model's prediction. As shown in the formula, we balance the probabilities of entities and candidate pairs by taking the square root, which keeps the probability within the range of two related probabilities. This approach can avoid unilateral decrease of probability and better meeting the expectation of the model.

$$p(e) = \sqrt{p\left(e_{\text{start}}\right) * p\left(e_{end}\right)} \tag{5}$$

$$p(a, o) = \begin{cases} \sqrt{p(a) * p(o \mid a)} \cdots \text{ if } (a, o) \in V_{AO} \\ \sqrt{p(o) * p(a \mid o)} \cdots \text{ if } (a, o) \in V_{OA} \end{cases} \tag{6}$$

where e represents the aspect or opinion entity, $start$ and end represent the start and end positions of the entity, and $p(a, o)$ represents the probability of the final candidate pair.

Finally, we employ the dynamic sentiment query $q_i^{D'}$ to predict the various aspects of emotion a_i. We obtain the output of labeled triplets for input sentence X_i, denoted as $T_i = \{(a, o, s)\}$, where $s \in$(positive, neutral, negative) and (a, o, s) refers to (aspect term, opinion term, sentiment polarity).

3.7 Loss Function Construction

In order to learn triplet subtasks jointly and make them promote each other, we integrate loss functions from various queries. For static queries in different directions, we minimize the loss of cross-entropy:

$$L_S = -\sum_{i=1}^{|Q^S|} \sum_{j=1}^{|S|} \left[p_{x_j,q_i}^{start} \cdot \log \hat{p}_{x_j,q_i}^{start} + p_{x_j,q_i}^{end} \cdot \log \hat{p}_{x_j,q_i}^{end}\right] \tag{7}$$

where $p()$ represents the distribution of gold, $\hat{p}()$ indicates the predicted distribution.

Similarly, the loss of dynamic queries in different directions is as follows:

$$L_D = -\sum_{i=1}^{|Q^D|} \sum_{j=1}^{|D|} \left[p_{x_j,q_i}^{start} \cdot \log \hat{p}_{x_j,q_i}^{start} + p_{x_j,q_i}^{end} \cdot \log \hat{p}_{x_j,q_i}^{end}\right] \tag{8}$$

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

Datasets	Train		Dev		Test	
	#S	#T	#S	#T	#S	#T
14-Lap	920	1265	228	337	339	490
14-Res	1300	2145	323	524	496	862
15-Res	593	593	148	238	318	455
16-Res	842	1289	210	316	320	465

Table 1: Statistics of 4 datasets. # S and # T denotes number of sentences and triples.

For dynamic sentiment classification queries, we minimize the cross-entropy loss function:

$$L_{D'} = -\sum_{i=1}^{\left|Q^{D'}\right|} p_{X,q_i}^{D'} \cdot \log \hat{p}_{X,q_i}^{D'} \tag{9}$$

Then, we integrate the aforementioned loss functions to generate the overall model's losses. In this paper, we used the method of AdamW (Loshchilov and Hutter, 2017) to optimize:

$$L(\theta) = L_S + L_D + L_{D'} \tag{10}$$

4 Experiments

4.1 Datasets

To verify the validity of our proposed approach, we conducted experiments on four benchmark datasets from the SemEval ABSA challenge (Pontiki et al., 2014; Pontiki et al., 2015; Pontiki et al., 2016) and listed the statistics for these datasets in Table 1.

4.2 Subtasks and Baselines

To demonstrate the validity of the proposed model, we compared the R-MMRC with the following baseline.

- CMLA+ (Peng et al., 2020) modifies CMLA (Yu et al., 2021), the attention mechanism is used by CMLA to detect the relationship between words and to extract aspects and opinions jointly. CMLA+ incorporates MLP to further determine whether the triplet is accurate during the matching phase.

- Two-Stage (Peng et al., 2020) is a two-stage pipeline model for ASTE. The task of the first stage is to mark all aspects and opinions. The goal of the second stage is to match all aspects with the corresponding opinion expression.

- RACL+ is improved by RACL framework (Chen and Qian, 2020), which uses mechanisms for relationship propagation and multi-task learning to enable subtasks to cooperate in a stacked multi-layer network. Then researchers (Chen et al., 2021) construct the query "Matching aspect a_i and opinion expression o_j?" to detect relationships.

- JET (Xu et al., 2020b) is a first end-to-end model with a novel position-aware tagging scheme that is capable of jointly extracting the triple.

- GTS-BERT (Wu et al., 2020) address the ASTE task in an end-to-end fashion with one unified grid tagging task.

- BMRC (Chen et al., 2021) transforms the ASTE task into a bi-directional MRC task and designs three types of queries to establish relationships between different subtasks.

	Models	14Lap			14Res			15Res			16Res		
		AESC	Pair	ASTE	AESC	Pair	ASTE	AESC	Pair	ASTE	AESC	Pair	ASTE
Precision	CMLA+	54.70	42.10	31.40	67.80	45.17	40.11	49.90	42.70	34.40	58.90	52.50	43.60
	TS	63.15	50.00	40.40	74.41	47.76	44.18	67.65	49.22	40.97	71.18	52.35	46.76
	RACL+	59.75	54.22	41.99	75.57	73.58	62.64	68.35	67.89	55.45	68.53	72.77	60.78
	JET	-	-	52.00	-	-	66.76	-	-	59.77	-	-	63.59
	GTS-BERT	-	66.41	57.52	-	76.23	70.92	-	66.40	59.29	-	71.70	68.58
	BMRC	**72.73**	74.11	**65.12**	77.74	76.91	71.32	72.41	**71.59**	63.71	**73.69**	76.08	67.74
	Ours	70.32	**74.60**	63.76	**78.95**	**78.36**	**72.69**	**72.95**	69.57	**63.96**	72.22	**78.04**	**68.64**
Recall	CMLA+	59.20	46.30	34.60	73.69	53.42	46.63	58.00	46.70	37.60	63.60	47.90	39.80
	TS	61.55	58.47	47.24	73.97	68.10	62.99	64.02	65.70	54.68	72.30	70.50	62.97
	RACL+	**68.90**	**66.94**	51.84	**82.23**	67.87	57.77	**70.72**	63.74	52.53	**78.52**	71.83	60.00
	JET	-	-	35.91	-	-	49.09	-	-	42.27	-	-	50.97
	GTS-BERT	-	64.95	51.92	-	74.84	69.49	-	68.71	58.07	-	**77.79**	66.60
	BMRC	62.59	61.92	54.41	75.10	75.59	70.09	62.63	65.89	58.63	72.69	76.99	**68.56**
	Ours	62.92	63.27	**54.69**	77.00	**78.54**	**72.85**	68.49	**70.33**	**62.64**	68.49	70.33	67.31
F1-score	CMLA+	56.90	44.10	32.90	70.62	48.95	43.12	53.60	44.60	35.90	61.20	50.00	41.60
	TS	62.34	53.85	43.50	74.19	56.10	51.89	65.79	56.23	46.79	71.73	60.04	53.62
	RACL+	64.00	59.90	46.39	**78.76**	70.61	60.11	69.51	65.46	53.95	**73.19**	72.29	60.39
	JET	-	-	42.48	-	-	56.58	-	-	49.52	-	-	56.59
	GTS-BERT	-	65.67	54.58	-	75.53	70.20	-	67.53	58.67	-	74.62	67.58
	BMRC	**67.27**	67.45	59.27	76.39	76.23	70.69	67.16	68.60	61.05	73.18	76.52	68.13
	Ours	66.41	**67.61**	**61.45**	77.96	**78.45**	**72.77**	**69.70**	**69.95**	**62.30**	72.41	**77.62**	**69.67**

Table 2: Statistics of 4 datasets. # S and # T denotes number of sentences and triples.

4.3 Model Settings and Evaluation Metrics

We adopted a Bert (Xu et al., 2019) model for the encoding layer with 12 attention heads, 12 hidden layers, and 768 hidden sizes. The fine-tuning rate of BERT and the learning rate of the training classifier are set to 1e-5 and 1e-3, respectively. We use AdamW optimizer with a weight decay of 0.01 and a warm-up rate of 0.1. At the same time, we set the batch size to 8 and the dropout rate to 0.3. The F1-score is extracted according to the triplet state on the development set. The threshold λ manually adjusted to 0.8, and the step size is set to 0.1.

We use precision, recall, and f1-score as measurement indicators to measure performance, including aspect term and sentiment co-extraction, aspect-opinion pair extraction, and aspect sentiment triplet extraction, respectively. Only when the prediction of aspects, opinions, and sentiments is correct, the triplet's prediction is correct.

4.4 Main Results

Table 2 shows the comparison results for all approaches, from which we derive the following conclusions. The proposed model R-MMRC achieves competitive performance on all datasets, which demonstrates the efficacy of our model. Under the F1 metric, the R-MMRC model is superior to the pipeline method in all datasets. Our model's F1-score on AESC exceeded the baseline average by 2.09%, on Pair by 3.66%, and on ASTE by 2.67%, respectively. The result shows that our method extracts more practical features. We observe that the method based on MRC achieves more significant improvement than the pipeline method, because it establishes the correlation between these subtasks by jointly training multiple subtasks, and alleviates the error propagation problem. It is worth noting that our model also has a significant improvement in precision, which indicates that the model's prediction ability is more reliable than those baselines.

The Pair and ASTE of our model achieve the best performance on all datasets, but the scores of two datasets in AESC are inferior to RACL+. We think that the idea that RACL+ first jointly trains the underlying shared features, then independently trains the advanced private features, and finally exchanges subtask information clues through the relationship propagation mechanism is very effective. TS performs better than CMLA+, since it uses a unified tagging schema to resolve sentiment conflicts. It is noteworthy that the improvement of precision contributes the most to the increase in F1 score. We believe that the

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

Model	14Lap			14Res			15Res			16Res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
R-MMRC	63.76	**54.69**	**61.45**	**72.69**	**72.85**	**72.77**	**63.96**	**62.64**	**62.30**	**68.64**	67.31	**69.67**
—rethink mechanism	**64.45**	53.21	58.30	71.76	65.42	68.34	60.21	59.26	59.57	67.61	65.02	67.32
—exclusive classifier	63.60	55.26	60.58	72.02	68.91	72.36	63.67	61.85	61.98	68.50	**68.39**	69.15
—probability generation	62.51	53.03	59.03	70.64	69.73	70.50	61.16	60.03	60.69	67.26	66.16	67.80
—dynamic query	60.12	50.41	53.27	65.32	67.63	61.16	55.71	56.63	54.05	62.74	60.56	60.13

Table 3: Ablation study results (%). P represents precision, R represents recall, F1 represents Macro-F1 score.

high precision score is due to the rethink mechanism filtering out some negative samples. Both JET and GTS-BERT used labeling schemes, but the latter yielded better results due to the use of more advanced grid labeling and the design of effective inference strategies. The sentiment classification task is more challenging than the previous extraction task because sentiment heavily relies on the extracted aspect-opinion pairs. However, with the help of dynamic sentiment queries constructed based on aspect-opinion information, compared to BMRC, an overall improvement has been achieved.

There is a certain performance gap between the baseline model and our proposed model, which confirms the rationality of the architecture we proposed. We believe that the design of static and dynamic queries can naturally integrate entity extraction and relation detection to enhance their dependency. The rethink mechanism validates each candidate aspect-opinion pair by modeling the information flow from aspect to opinion (or from opinion to aspect), effectively filtering out negative samples and improving the performance of the model. At the same time, the exclusive classifier we introduced, as well as the probability generation algorithm, further improve the performance of the model.

4.5 Ablation Test

We conduct further ablation studies to analyze the impact of different components of R-MMRC. We present the results of ASTE in Table 3, where the first row shows the reproduced results of R-MMRC. The next three rows show the results after removing the rethink mechanism, exclusive classifier, and probability generation, respectively. The last row shows the final results after removing these three parts of the R-MMRC model.

The results show that each component improves the performance of the model, demonstrating their advantages and effectiveness. We remove the dynamic query in the parameter sharing stage of R-MMRC and keep only static queries and the dynamic sentiment query, which is referred to as "-dynamic query". Obviously, removing the dynamic query resulted in a significant drop in model performance. We analyze that after removing the dynamic query, the model could not capture the dependency relationships between entities and separated entity extraction from relation detection. The results indicate that the dynamic query in the parameter sharing stage is highly effective in capturing dependencies.

The advantage of the rethink mechanism is quite significant. Specifically, compared with R-MMRC, the rethink mechanism achieved F1-score improvements of 3.15%, 3.43%, 2.73%, and 2.35% on the four datasets, demonstrating the effectiveness of the rethink mechanism. The probability generation also has a certain improvement effect, which proves that our model better avoids unilateral decline of probability and is more consistent with the model's expectation. For the exclusive classifier, the model's F1 score improvement is relatively smaller compared to the previous two components. Moreover, we find that it has a significant downside of slowing down the model's runtime.

4.6 Case Study

We conduct a case study to illustrate the effectiveness and perform an error analysis in Table 4. We select three cases from datasets and compare our results with RACL+. The reason for choosing RACL+ is that its performance is second only to our R-MMRC model.

The first case has two aspect terms: "exterior patio" and "ambiance". RACL+ cannot extract the triplets corresponding to "ambiance". We speculate that the model only considers the relationship be-

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

Case	Ground Truth	RACL+	R-MMRC
The outdoor patio is really nice in good weather, but what ambience the indoors possesses is negated by the noise and the crowds.	(outdoor patio, nice, POS) (ambience, negated, NEG)	(outdoor patio, nice, POS)	(outdoor patio, nice, POS) (ambience, negated, NEG)
The food is pretty good, but after 2 or 3 bad experiences at the restaurant (consiste--ntly rude, late with RSVP'd seating).	(food, pretty good, POS) (seating, RSVP, NEU)	(food, pretty good, POS) (seating, rude, NEG) × (seating, late, NEG) ×	(food, pretty good, POS) (seating, RSVP, NEU)
Dinner is okay not many vegetarian options and the portions are small.	(Dinner, okay, NEU) (positions, small, NEG)	(Dinner, okay, POS) × (vegetarian options, not many, NEG) × (portions, small, NEG)	(Dinner, okay, POS) × (vegetarian options, not many, NEG) × (portions, small, NEG)

Table 4: Case study. Marker × indicates incorrect predictions. The table's abbreviations POS, NEU, and NEG represent positive, neutral, and negative sentiments, respectively.

tween sentence representations of subtasks, which weakens aspect terms in long and complicated sentences. Our proposed model considers all triplets in the sentence because it can guarantee that an aspect or an opinion can produce a pair, precisely like human reading behavior.

The second case is a long sentence with two triplets, and the corresponding sentiments are positive and neutral, respectively. Our R-MMRC correctly extracted aspect terms and opinion terms, and successfully predicted the corresponding polarity. However, RACL+ correctly extracts all aspect terms, but it misjudges the polarity of "seating". The reason is that RACL+ is good at making use of different semantic relationships between subtasks, so it may use irrelevant "rule" and "late" as keywords, and predict the sentiment of "seating" as "negative". On the contrary, R-MMRC can more accurately identify aspect terms and the corresponding opinion terms in complex sentences.

The third case is error analysis. Although the sentence is not long, both models predict the sentiment of "dinner" incorrectly. We analyze that "ok" is usually considered a positive opinion term, so the two models define "dinner" as positive. However, by carefully observing this sentence, we find that the seldom choices in "vegetarian options " are the reason why guests say "dinner" is just "okay" rather than "good". So, sentiment polarity should be "neutral" rather than "positive". We speculate that we are looking for the training loss of maximum likelihood cross entropy in the training set, which may be the reason for the wrong prediction in this case. More interestingly, RACL+ and our R-MMRC, as two excellent solutions, incorrectly consider (vegetarian options, not many, NEG) as a triplet. Therefore, we think that understanding sentence structure through logic and even causal reasoning may provide new ideas for the future research of sentiment analysis.

5 Conclusion

In this paper , we investigate ASTE task and propose an improved multi-round MRC framework with a rethink mechanism(R-MMRC). This framework sequentially extracts aspect-sentiment pairs and performs sentiment classification, which can handle complex correspondences between aspects, opinions, and sentiments. In each round, explicit semantic information can be effectively utilized. Additionally, the rethink mechanism models the bidirectional information flow to verify each candidate aspect-opinion pair, effectively utilizing the corresponding relationship between entities. Exclusive classifiers avoid interference between different queries, and probability generation algorithms further improve prediction performance. The experimental results demonstrate the effectiveness of the R-MMRC framework, further improving the overall performance of the system. More importantly, our model can serve as a general framework to address various tasks of ABSA. However, our model still suffers from the issue of high computational cost, and we hope to compress the model in the future to make it more lightweight.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.

Acknowledgements

This work is supported by a grant from the Social and Science Foundation of Liaoning Province (No. L20BTQ008)

References

Zhuang Chen and Tieyun 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.

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 12666–12674.

Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2974–2985.

Zhifang Fan, Zhen Wu, Xinyu 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.

Lei Gao, Yulong Wang, Tongcun Liu, Jingyu Wang, Lei Zhang, and Jianxin Liao. 2021. Question-driven span labeling model for aspect–opinion pair extraction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 12875–12883.

Devamanyu Hazarika, Soujanya Poria, Prateek Vij, Gangeshwar Krishnamurthy, Erik Cambria, and Roger Zimmermann. 2018. Modeling inter-aspect dependencies for aspect-based sentiment analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 266–270.

Binxuan Huang and Kathleen M Carley. 2019. Parameterized convolutional neural networks for aspect level sentiment classification. *arXiv preprint arXiv:1909.06276*.

Ilya Loshchilov and Frank Hutter. 2017. Fixing weight decay regularization in adam.

Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 2019. Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics*, pages 3538–3547.

Navonil Majumder, Soujanya Poria, Alexander Gelbukh, Md Shad Akhtar, Erik Cambria, and Asif Ekbal. 2018. Iarm: Inter-aspect relation modeling with memory networks in aspect-based sentiment analysis. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3402–3411.

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.

M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, and S. Manandhar. 2014. Semeval-2014 task 4: Aspect based sentiment analysis. *Proceedings of International Workshop on Semantic Evaluation at*.

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

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammed AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, et al. 2016. Semeval-2016 task 5: Aspect based sentiment analysis. In *ProWorkshop on Semantic Evaluation (SemEval-2016)*, pages 19–30. Association for Computational Linguistics.

Qianlong Wang, Zhiyuan Wen, Qin Zhao, Min Yang, and Ruifeng Xu. 2021. Progressive self-training with discriminator for aspect term extraction. In *Proceedings of the 2021 conference on empirical methods in natural language processing*, pages 257–268.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. *arXiv preprint arXiv:2010.04640*.

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

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

Lu Xu, Lidong Bing, Wei Lu, and Fei Huang. 2020a. Aspect sentiment classification with aspect-specific opinion spans. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3561–3567.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020b. Position-aware tagging for aspect sentiment triplet extraction. In *Conference on Empirical Methods in Natural Language Processing*.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. *arXiv preprint arXiv:2107.12214*.

Hang Yan, Junqi Dai, Xipeng Qiu, Zheng Zhang, et al. 2021. A unified generative framework for aspect-based sentiment analysis. *arXiv preprint arXiv:2106.04300*.

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.

Mi Zhang and Tieyun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*, pages 3540–3549.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2022. A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. 2020. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*, pages 3239–3248.

Yukun Zheng, Jiaxin Mao, Yiqun Liu, Zixin Ye, Min Zhang, and Shaoping Ma. 2019. Human behavior inspired machine reading comprehension. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 425–434.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 701-712, Harbin, China, August 3 - 5, 2023.