

Self Question-answering: Aspect-based Sentiment Analysis by Role Flipped Machine Reading Comprehension

Guoxin Yu[✉], Jiwei Li[✉], Ling Luo[✉], Yuxian Meng[✉], Xiang Ao[✉], Qing He[✉]

[✉]Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing 100190, China.

[✉]University of Chinese Academy of Sciences, Beijing 100049, China.

[✉]Institute of Intelligent Computing Technology, Suzhou, CAS

[✉]Shannon.AI, China.

{yuguoxin20g, luoling18s, aoxiang, heqing}@ict.ac.cn

{jiwei_li, yuxian_meng}@shannonai.com

Abstract

The pivot for the unified Aspect-based Sentiment Analysis (ABSA) is to couple aspect terms with their corresponding opinion terms, which might further burgeon easier sentiment predictions. In this paper, we investigate the unified ABSA task from the perspective of Machine Reading Comprehension (MRC) by observing that the aspect and the opinion terms can serve as the query and answer in MRC interchangeably. We propose a new paradigm named Role Flipped Machine Reading Comprehension (RF-MRC) to resolve. At its heart, the predicted results of either the Aspect Term Extraction (ATE) or the Opinion Terms Extraction (OTE) are regarded as the queries, respectively, and the matched opinion or aspect terms are extracted as answers. The queries and answers can be flipped for multi-hop detection. Finally, every matched aspect-opinion pair is predicted by the sentiment classifier. RF-MRC can solve the ABSA task without extra data annotation. Experiments on three widely used benchmarks and a challenging dataset demonstrate the superiority of the proposed framework.

1 Introduction

Aspect-based Sentiment Analysis (ABSA) aims at detecting opinions towards different targets (also known as aspects) instead of inferring overall sentiment polarity in a given sentence (Liu, 2012). It generally consists of three fundamental sub-tasks, namely, aspect terms extraction (ATE), opinion terms extraction (OTE), and aspect sentiment classification (ASC). ATE and OTE extract aspect and

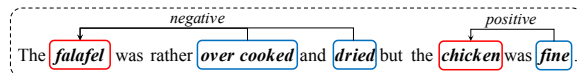


Figure 1: An illustrative example of the connection between aspect terms and opinion terms.

opinion terms from sentences, respectively. And ASC predicts the sentiment polarities (i.e., *positive*, *negative*, and *neutral*) towards aspect terms.

Practically, the heart of ASBA is to capture the *connection* between aspect terms and their respective opinion terms, which might make it easier to predict the correct sentiment polarities. Such connection is more substantial when multiple aspects with different polarities exist. For example, we illustrate the connection within a sentence shown in Figure 1. The *negative* polarity of “*falafel*” can be derived by an aggregation of the relevant opinions “*over cooked*” and “*dried*”, whereas the *positive* polarity of “*chicken*” is oriented by its corresponding opinion word “*fine*”. If the aspect terms and their connected opinion words are mismatched, the prediction may become difficult even incorrect.

Hence, immense efforts have been dedicated to grasping the relations between aspect terms and their potential corresponding opinion terms. Early methods only focused on ASC task and relied on given aspect terms. Among them, a series of methods designed attention mechanisms (He et al., 2018; Tang et al., 2019) or gating mechanisms (Zhang et al., 2016; Xue and Li, 2018) to collect aspect-related information (e.g., opinion terms) from context. Recently, Graph Neural Network over different dependency trees (Huang and Carley, 2019;

^{*}Corresponding author

Tang et al., 2020; Hou et al., 2021b) was proposed to link aspect terms with interrelated opinion terms more directly. They can account for long-range word dependencies and refrain from identifying contextual words unrelated to aspect terms.

Despite their effectiveness, these methods will be infeasible if the given aspects are absent. As a result, some researchers proposed to incorporate all sub-tasks in a framework of unified ABSA. These methods (He et al., 2019; Chen and Qian, 2020) formulated sub-tasks of ABSA as sequence labeling tasks. By multifarious interaction mechanisms performed on sentence representations of different sub-tasks, they made the aspect terms come into contact with opinion terms. Furthermore, recent researches (Peng et al., 2020; Mao et al., 2021) put forward to extract (*aspect, opinion, sentiment*) triples from sentences without given aspect terms. They strive to clarify each aspect-opinion pair for sentiment prediction and needed additional labels of triples compared to the previous unified ABSA.

In this paper, we examine the unified ABSA from a perspective of Machine Reading Comprehension (MRC). The MRC framework operates on the context, query, and answer triples (Rajpurkar et al., 2016, 2018), in which the constructed natural language query is asked to the context, and the answer is extracted from the context.

By observing that the aspect terms and opinion terms can be naturally characterized as queries and answers, we propose a new paradigm named Role Flipped Machine Reading Comprehension (RF-MRC) to meet the heart of the unified ABSA.

First, we extract the initial aspect and opinion terms from a given sentence. Then either the initial aspect terms or opinion terms are deemed as a query to extract corresponding opinion terms or aspect terms as answers. The roles of query and answer can be flipped to perform a multi-hop question-answering process. In this manner, we can progressively obtain the aspect or opinion terms we need without manually designing queries. Meanwhile, the aspect terms could be potentially associated with relevant opinion terms as the multiple question-answering proceeds. Furthermore, we propose a matching module to match all the extracted aspects and relevant opinion terms in pairs simultaneously instead of extracting only one aspect-opinion pair at one time, considering a complex sentence may contain multiple aspects with conflict polarities. Experiments on three widely

used benchmarks and a challenging dataset demonstrate the superiority of the proposed framework.

2 Related Work

2.1 Aspect-based Sentiment Analysis

Existing methods for ABSA consist of *separate learning* and *joint learning*, respectively. Methods for *separate learning* only focus on one of the sub-tasks of ABSA. To name some, Wang et al. (2016a); Li and Lam (2017); Angelidis and Lapata (2018); Ma et al. (2019); Li et al. (2020a); Luo et al. (2019b) came up with different un/supervised methods to solve aspect extraction. Tang et al. (2019); Liang et al. (2019); Du et al. (2019); Chen and Qian (2019); Tian et al. (2020); Huang et al. (2020); Xu et al. (2021) designed different neural networks with attention mechanisms to exploit contextual and positional proximity related to aspect terms for sentiment prediction. Sun et al. (2019); Zhang et al. (2019); Hou et al. (2021a); Wang et al. (2020); Tang et al. (2020) established graph neural network over dependency trees to capture long-range syntactic relations between aspect terms and relevant opinion terms.

Joint learning methods strive to solve multiple sub-tasks simultaneously. Hu et al. (2019); Phan and Ogunbona (2020) used pipeline models to extract aspect terms then predict the sentiment polarities, which are vulnerable due to error accumulation. To tackle this issue, some studies proposed to solve all sub-tasks in a joint learning framework. Wang et al. (2018); Li et al. (2019b) used a unified tagging schema to solve ATE and ASC simultaneously. Wang et al. (2017); Dai and Song (2019); Luo et al. (2019a); Chen et al. (2020); Zhao et al. (2020) integrated ATE and ASC in the same framework to make these two tasks benefit from each other. Some emerging methods (He et al., 2019; Chen and Qian, 2020; Peng et al., 2020; Xu et al., 2020; Yu et al., 2021) added OTE as an auxiliary task and connect aspects with respective opinion terms to derive easier sentiment prediction. In addition, recent studies defined a task of (*aspect, opinion, sentiment*) triples extraction and resolve it in a two-stage framework (Peng et al., 2020) or a unified framework (Mao et al., 2021). Nevertheless, this task demands supplementary data to mark precise (aspect, opinion, sentiment) triples.

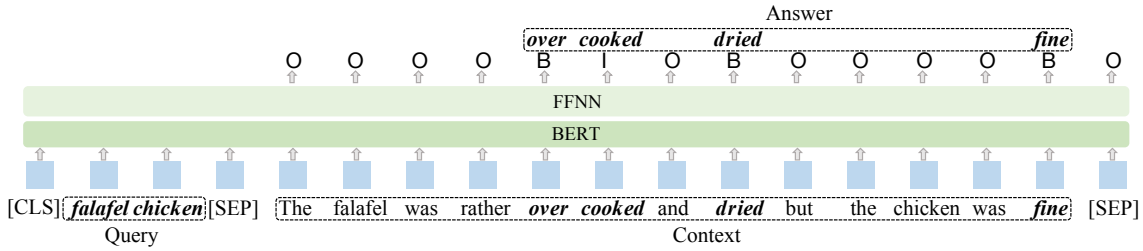


Figure 2: An example to examine the unified ABSA from a perspective of MRC.

2.2 Solving NLP Tasks by MRC

Machine reading comprehension is a prevalent and elastic framework, which aims to extract answers from context according to query. Many tasks in natural language processing can be framed as comprehension reading.

McCann et al. (2018) introduced a natural language decathlon and transformed ten tasks into reading comprehension problems. He et al. (2015) used question-answering pairs to represent the predicate-argument structure in the semantic role labeling annotations. Levy et al. (2017) showed that relation extraction can be reduced to answer simple reading comprehension questions. Li et al. (2020b) designed a unified machine reading comprehension framework to solve the task of nested named entity recognition. Li et al. (2019a) cast the entity-relation extraction as a multi-turn question answering problem. Wu et al. (2020) used a mention with its surrounding words as a query to extract its coreference words as answers. All the above methods have demonstrated that machine reading comprehension is an effective framework to solve natural language processing tasks.

3 Model

3.1 The formulation of unified ABSA

Given a sequence of tokens $X = \{x_1, x_2, \dots, x_n\}$, where n denotes the length of sentence, Aspect Terms Extraction (ATE) aims to find aspect terms in X and assign a label $\hat{Y}^A = \{\hat{y}_1^A, \hat{y}_2^A, \dots, \hat{y}_n^A\}$ to it. Opinion Terms Extraction (OTE) aims to find all opinion terms in X and assign a label of $\hat{Y}^O = \{\hat{y}_1^O, \hat{y}_2^O, \dots, \hat{y}_n^O\}$ to it. Aspect Sentiment Classification (ASC) aims to predict a sequence of sentiment label $\hat{Y}^S = \{\hat{y}_1^S, \hat{y}_2^S, \dots, \hat{y}_n^S\}$. Specifically, $\hat{y}_i^A, \hat{y}_i^O \in \{B, I, O\}$ denote the beginning of, inside of, and out of aspect and opinion terms, respectively. $\hat{y}_i^S \in \{pos, neg, neu\}$ denotes positive, negative, neutral sentiment polarities, respectively.

Sentiment labels of tokens that are not aspects are set to “NULL”.

3.2 Examine ABSA from MRC perspective

Recall that the Machine Reading Comprehension (MRC) aims to determine the answer to a given query from context. The query encodes significant prior information and the answer can be extracted by detecting its association with the query within context. This configuration provides an elegant way to capture the connection between aspect terms and relevant opinion terms.

In the light of such observation, we examine the unified ABSA from the perspective of MRC. The input sentence is naturally regarded as context. Then, the query could be constructed by aspect terms (opinion terms) and the answer consists of corresponding opinion terms (aspect terms) related to its query. Through this manner, aspect terms come into contact with corresponding opinion terms, and vice versa, by interactions between query and answer. In this way, we believe the unified ABSA can be solved by an MRC framework. For implementation, we can simply concatenate the query and context then feed them into BERT and a feed forward neural network to get the answer, which is exhibited in Figure 2.

In this paper, we proposed a paradigm named Role Flipped Machine Reading Comprehension (RF-MRC) to meet the heart of unified ABSA and derive easier sentiment classification. The overall architecture is shown in Figure 3 and the algorithm is elaborated in appendix of the supplementary materials.

3.3 Input Representations

We use BERT (Devlin et al., 2019) to obtain input representation following (Li et al., 2019b) and (Chen and Qian, 2020). For a sequence of tokens $X^{(0)} = \{x_1, x_2, \dots, x_n\}$, we map the word sequence with pre-trained BERT model to generate a sequence of units vectors $H^{(0)} =$

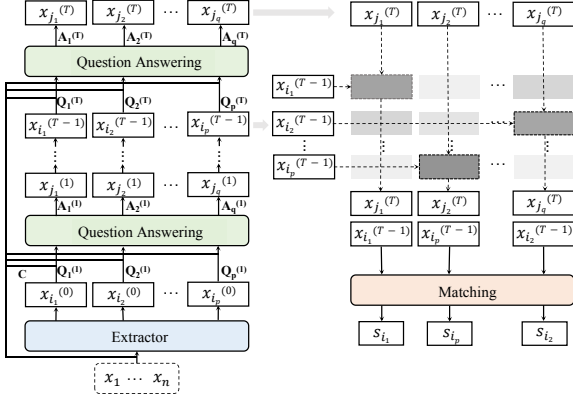


Figure 3: Model Overview

$\{h_1^{(0)}, h_2^{(0)}, \dots, h_n^{(0)}\} \in \mathbb{R}^{d_h \times n}$, where d_h denotes the dimension of word vectors.

3.4 Initial Terms Extraction

In this section, we extract p candidate aspect terms and q opinion terms from the initial sentence with blank queries.

As shown in Figure 3, we perform the initial extraction of aspect or opinion terms with a blank query. For the word vectors $H^{(0)} = \{h_1^{(0)}, h_2^{(0)}, \dots, h_n^{(0)}\}$, we first use a feed forward neural network to get the sequence labels for ATE:

$$\begin{aligned} (\hat{Y}^A)^{(0)} &= \text{FFNN}(H^{(0)}), \\ (\hat{Y}^O)^{(0)} &= \{(\hat{y}_1^A)^{(0)}, (\hat{y}_2^A)^{(0)}, \dots, (\hat{y}_n^A)^{(0)}\}, \end{aligned} \quad (1)$$

where $^{(0)}$ denotes the initial question answering process with a blank query, FFNN denotes feed forward neural network. We select the top p candidate aspect terms $(X^A)^{(0)}$ from X ,

$$(X^A)^{(0)} = \{(x_{i_1})^{(0)}, (x_{i_2})^{(0)}, \dots, (x_{i_p})^{(0)}\}. \quad (2)$$

where i_i denotes the indexes of top p potential aspect terms in the sentence. Similarly, we could get the sequence labels $(\hat{Y}^O)^{(0)} = \{(\hat{y}_1^O)^{(0)}, (\hat{y}_2^O)^{(0)}, \dots, (\hat{y}_n^O)^{(0)}\}$ for OTE based on the word vectors $H^{(0)}$. Then, the top q candidate opinion terms $(X^O)^{(0)} = \{(x_{j_1})^{(0)}, (x_{j_2})^{(0)}, \dots, (x_{j_q})^{(0)}\}$ are extracted from X . Note that j_j denotes the indexes of top q potential opinion terms in the sentence.

3.5 Role Flipped Module

Based on the initial extraction results, we devise a role flipped module to grasp the connection between aspect terms and relevant opinion terms inside the sentence. The process is shown as the left

part of the Figure 3. First, given the sentence as the context, we take the extracted aspect terms as queries to extract corresponding opinion words as answers. The queries are constructed by $(X^A)^{(0)}$ and the context is the input sentence $X^{(0)}$. In this round, the input can be formed as follows:

$$\begin{aligned} (X)^{(1)} &= \{[\text{CLS}], (x_{i_1})^{(0)}, \dots, (x_{i_p})^{(0)}, [\text{SEP}], \\ &\quad x_1, \dots, x_n, [\text{SEP}]\}. \end{aligned} \quad (3)$$

We feed it into BERT to get hidden vectors $H^{(1)}$. Then a feed forward neural network is used to get the labels of opinion terms as answers:

$$(\hat{Y}^O)^{(1)} = \text{FFNN}(H^{(1)}), \quad (4)$$

where FFNN denotes feed forward neural network and $^{(1)}$ represents the hop number. Then we flip the query and the answer for the next round of question-answering. The above process can be iterated into a multi-hop question-answering process. Noted that answers in the t -th round will serve as queries in the $t+1$ round. After T rounds of question answering processes, we get the final aspect terms $(X^A)^{(T-1)}$ and the opinion terms $(X^O)^{(T)}$ based the labels $(\hat{Y}^A)^{(T-1)}$ and $(\hat{Y}^O)^{(T)}$ in the last round:

$$\begin{aligned} (X^A)^{(T-1)} &= \{(x_{i_1})^{(T-1)}, \dots, (x_{i_p})^{(T-1)}\}, \\ (X^O)^{(T)} &= \{(x_{j_1})^{(T)}, \dots, (x_{j_q})^{(T)}\}. \end{aligned} \quad (5)$$

Specifically, we set the aspect terms as queries and the opinion terms as answers in the last round.

Analogously, we can first take the extracted opinion terms $(X^O)^{(0)}$ in the initial terms extraction as queries. Then the same multi-hop question-answering process is performed to get the final opinion terms $(X^O)^{(T-1)}$ and aspect terms $(X^A)^{(T)}$ after T rounds. For convenience, we call the process where aspect terms are firstly taken as queries as ‘‘A2O’’, and the other is called ‘‘O2A’’.

3.6 Matching Module

So far, we have extracted all the aspect terms and corresponding opinion terms. In order to exploit the captured connection between them, we propose a matching mechanism to match them in pairs and derive easier sentiment prediction. For A2O, after T rounds of cross question answering, we get a set of candidate aspect terms $(X^A)^{(T-1)}$ and a set of opinion terms $(X^O)^{(T)}$. We apply an attention mechanism to compute the correspondence

between them:

$$\begin{aligned} score_{ij}^{(i \neq j)} &= (\mathbf{H}_i^{(T)})^T \mathbf{H}_j^{(T)}, \\ \mathbf{A}_{ij} &= \frac{\exp(score_{ij})}{\sum_{k=1}^n \exp(score_{ik})}, \end{aligned} \quad (6)$$

where $\mathbf{H}^{(T)}$ is hidden features in the last round and it encapsulates the captured connections in the role flipped module. We can select a best opinion term $(x_j)^T$ from $(\mathbf{X}^O)^{(T)}$ for each aspect term $(x_i)^{T-1} \in (\mathbf{X}^A)^{(T-1)}$, according to \mathbf{A}_{ij} . We use the word vectors of $(x_i)^{T-1}$ and $(x_j)^T$ to compute the sentiment scores of the aspect terms:

$$(\hat{y}_i^S)^{(1)} = \text{FFNN}((\mathbf{H}_i)^{T-1} : (\mathbf{H}_j)^T), \quad (7)$$

where “:” represents concatenation. Similarly, for O2A, we can take use of opinion terms $(\mathbf{X}^O)^{(T-1)}$ and aspect terms $(\mathbf{X}^A)^{(T)}$ in the last round to compute sentiment scores:

$$(\hat{y}_i^S)^{(2)} = \text{FFNN}((\mathbf{H}_i)^T : (\mathbf{H}_j)^{T-1}). \quad (8)$$

For a candidate aspect term x_i , the corresponding sentiment score is an average score:

$$\hat{y}_i^S = \frac{1}{2}((\hat{y}_i^S)^{(1)} + (\hat{y}_i^S)^{(2)}), i = 1, 2, \dots, n. \quad (9)$$

Here we only calculate the sentiment scores of aspect terms, the label \hat{y}_i^S for any other word is set to “NULL”.

In this manner, we can deploy all the extracted connections inside the sentence at once, without using auxiliary labels of triples like (Peng et al., 2020; Mao et al., 2021).

3.7 Training

Referring to the Figure 3, in every round of question answering, including the initial extraction, there are two predicted results of aspect terms and opinion terms. Suppose in the t -th round, the predicted labels are $(\mathbf{Y}^A)^{(t)}$ and $(\mathbf{Y}^O)^{(t)}$, for ATE and OTE respectively. Then we use the cross-entropy to compute the losses of ATE and OTE in the t round:

$$\begin{aligned} (\mathcal{L}^A)^{(t)} &= -\frac{1}{N} \sum_{i=1}^N \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij}^A \cdot \log(\hat{y}_{ij}^A)), \\ (\mathcal{L}^O)^{(t)} &= -\frac{1}{N} \sum_{i=1}^N \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij}^O \cdot \log(\hat{y}_{ij}^O)), \end{aligned} \quad (10)$$

where N denotes the number of training instances, n_i denotes the number of tokens in the i -th instance.

Dataset	Sentence	Aspect			Opinion	
		Pos	Neu	Neg		
Restaurant14	Train	3,044	2,164	807	637	3,484
	Test	800	728	196	196	1,008
Laptop14	Train	3,048	994	870	464	2,504
	Test	800	341	128	169	674
Restaurant15	Train	1,315	902	34	252	1,210
	Test	685	319	27	179	510
MAMS19	Train	4,297	3,380	5,042	2,764	-
	Dev	500	403	604	325	-
	Test	500	400	607	329	-

Table 1: The statistics of datasets.

After T rounds of question answering, the losses of ATE and OTE are as follows:

$$\begin{aligned} \mathcal{L}^A &= \sum_{t=0}^T \lambda_t^A \cdot (\mathcal{L}^A)^{(t)}, \\ \mathcal{L}^O &= \sum_{t=0}^T \lambda_t^O \cdot (\mathcal{L}^O)^{(t)}, \end{aligned} \quad (11)$$

λ_t^A and λ_t^O are coefficients of ATE and OTE in the t -th round. And in the last round, we get the final sentiment label $(\mathbf{Y}^S)^{(T)}$. We also use the cross-entropy to get the loss of ASC:

$$\mathcal{L}^S = -\frac{1}{N} \sum_{i=1}^N \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{ij}^S \cdot \log(\hat{y}_{ij}^S)). \quad (12)$$

where N denotes the number of training instances, n_i denotes the number of tokens in the i -th instance. The overall loss is the weighted sum of the sub-tasks’ losses:

$$\mathcal{L} = \alpha \cdot \mathcal{L}^A + \beta \cdot \mathcal{L}^O + \gamma \cdot \mathcal{L}^S. \quad (13)$$

α, β, γ are task coefficients.

4 Experiment

4.1 Datasets

We adopt three widely used datasets: Restaurant14 and Laptop14 from SemEval 2014 Task 4 (Pontiki et al., 2014), Restaurant15 from Semeval 2015 Task 12 (Pontiki et al., 2015). Note that these three datasets originally contain aspect term labels and sentiment labels, and labels for opinion terms are annotated by (Wang et al., 2016b). We also use a challenging dataset MAMS constructed by (Jiang et al., 2019), in which each sentence contains at least two aspects with different polarities, to perform comprehensive investigations. There are no opinion labels in MAMS. The forms of all datasets

are consistent with the description in 3.1 and the statistics are exported in Table 1. For Restaurant14, Laptop14, and Restaurant15, we randomly sample 20% of the training set as the validation set. While the original MAMS dataset contains the training, validation, and test sets.

4.2 Compared Methods

Pipeline Model. Following Chen and Qian (2020), we perform DECNN (Xu et al., 2018) and CMLA (Xu et al., 2018) for ATE, TNet (Li et al., 2018) and TCaps (Chen and Qian, 2019) for ASC to four pipeline models. SPAN (Hu et al., 2019) performed a multi-target extractor for ATE and designed a sentiment polarities classifier for ASC.

Unified Model. MNN (Wang et al., 2018) and E2E-ABSA (Li et al., 2019b) jointly solve ATE and ASC by using collapsed tagging schema. DOER (Luo et al., 2019a) used a dual cross-shared RNN mechanism to share information between different sub-tasks. IMN (He et al., 2019) is an interactive multi-task model for ATE and ASC, while OTE is confused into ATE. RACL (Chen and Qian, 2020) is a joint learning framework which can solve ATE, OTE and ASC jointly and exploit four relations between different sub-tasks.

Our model only needs three annotation sequences related to three sub-tasks, while Peng et al. (2020); Mao et al. (2021) demand several labels of (aspect, opinion, sentiment) triple for each sentence. For this reason, we did not involve them in our compared models.

4.3 Settings

We used the pre-trained BERT_{large} model to generate word vectors with $d_h=1024$. We set the number of multiple rounds, the number of candidate aspect terms p , and the number of candidate opinion terms q as 2, 8, and 5, individually. Since a word can be broken into multiple tokens with the BERT model, p and q are bigger than the true number of aspect terms and opinion terms. We trained the model for 80 epochs using Adam optimizer (Kingma and Ba, 2015) with a learning rate of $1e-5$ and batch size 8. The task coefficients $\{\lambda_t^A, \lambda_t^O, \alpha, \beta, \gamma\}$ are set to $\{1, 1, 1, 1, 1\}$. The code is implemented in PyTorch 1.9.0 and launched on Ubuntu server with a NVidia Tesla V100(32GB).

Following the protocols in He et al. (2019), we use four metrics, i.e., AE-F1, OE-F1, AS-F1, and Overall-F1, representing macro F1 scores for ATE,

OTE, ASC, and overall performance for complete ABSA. For an aspect containing multiple tokens, we take the polarity of the first token as the final ASC result. As for the Overall-F1, we take the result as correct only when both ATE and ASC results are correct. The metrics of the comparison method are calculated in the same way. The model achieving the best Overall-F1 on the validation sets is used for evaluation on the test set.

4.4 Main Results

In order to make a fair validation for the proposed model, we first compare our method with all the baseline models on Restaurant14, Laptop14, and Restaurant15, which are the most widely-used benchmarks for ABSA. Table 2 demonstrates the main results.

We have several observations from Table 2. Firstly, the unified models perform better than the pipeline models, which proves the effectiveness of exploiting the connections between sub-tasks. Secondly, RACL is a strong baseline model compared with IMN and SPAN because RACL takes the relations between ATE and OTE into consideration.

Thirdly, our proposed model achieves the best or second best performance compared with all the baseline models on different sub-tasks. On the one hand, the AE-F1 and OE-F1 are higher than most baseline models. We deduce this is because the extraction results in the last round of question answering can be modified by results in the current round. On the other hand, the sentiment prediction of RF-MRC is more accurate. Especially, RF-MRC achieves 1.45%, 1.91% and 1.81% improvements over the strongest baseline on the Overall-F1 of three datasets. The results prove that using the proposed RF-MRC can exploit the relations between aspect and opinion terms at a more fine-grained level, while other baseline models only consider relations between sentence representations of sub-tasks. More specifically, aspect terms and corresponding opinion terms will be paired owing to the interaction between query and answer in the role flipped module. Consequently, the sentiment prediction becomes more accurate based on these terms in pairs.

4.5 Auxiliary Experiments

To demonstrate the ability of the proposed model to analyze the sentiment in complex sentences, we run an auxiliary experiment on a more challenging MAMS (Jiang et al., 2019) dataset. Each sentence

Model	Restaurant14				Laptop14				Restaurant15				
	AE-F1	OE-F1	AS-F1	Overall-F1	AE-F1	OE-F1	AS-F1	Overall-F1	AE-F1	OE-F1	AS-F1	Overall-F1	
M ₁	CMLA+TNet	81.91	83.84	69.69	64.49	77.49	76.06	68.30	55.94	67.73	70.56	62.27	55.00
M ₂	CMLA+TCap	81.91	83.84	71.32	65.68	77.49	76.06	69.49	56.30	67.73	70.56	63.32	55.47
M ₃	DECNN+TNet	82.79	-	70.45	65.80	79.38	-	68.69	57.39	68.52	-	62.41	55.69
M ₄	DECNN+TCap	82.79	-	71.77	66.84	79.38	-	69.61	57.71	68.52	-	63.60	56.22
M ₅	MNN	83.05	84.55	68.45	63.87	76.94	77.77	65.98	53.80	70.24	69.38	57.90	56.57
M ₆	E2E-TBSA	83.92	84.97	68.38	66.60	77.34	76.62	68.24	55.88	69.40	71.43	58.81	57.38
M ₇	DOER	84.63	-	64.50	68.55	80.21	-	60.18	56.71	67.47	-	36.76	50.31
M ₈	SPAN	86.71	-	71.75	73.68	82.34	-	62.50	61.25	74.63	-	50.28	62.29
M ₉	IMN	84.06	85.10	75.67	70.72	77.55	81.00	<u>75.56</u>	61.73	69.90	73.29	70.10	60.22
M ₁₀	RACL	86.38	87.18	81.61	<u>75.42</u>	81.79	79.72	73.91	63.40	73.99	<u>76.00</u>	<u>74.91</u>	66.05
M ₁₁	RF-MRC	88.22	<u>86.62</u>	<u>81.28</u>	76.87	82.44	<u>80.52</u>	76.05	65.31	75.57	78.60	75.79	67.86

Table 2: Comparison results. The best scores are in bold face and the second best ones are underlined. The scores for models from M₁ to M₁₀ are taken from Chen and Qian (2020). Models from M₁ to M₇ are based double embeddings (Xu et al., 2018), while M₈ to M₁₁ used BERT_{large} as a backbone. ‘-’ denotes the method does not have the metric OE-F1.

Model	AE-F1	AS-F1	Overall-F1
SPAN	73.90	82.51	61.51
IMN	73.03	84.29	61.68
RACL	75.14	83.63	63.03
RF-MRC	76.00	84.71	64.53

Table 3: Auxiliary results in MAMS.

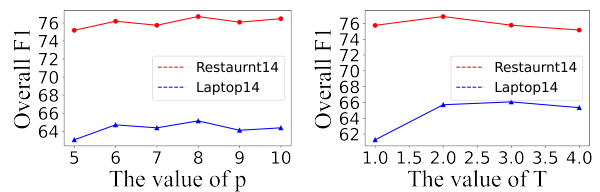
Model	Restaurant14	Laptop14	Restaurant15
w/o A2O	74.01	63.62	67.77
w/o O2A	75.21	64.22	67.38
Full Model	76.87	65.31	67.86

Table 4: Ablation Test. ‘w/o’ denotes without.

in this dataset consists of at least two unique aspects with different polarities. Because the opinion labels are not annotated in MAMS, we did not compute the loss \mathcal{L}^O and only use three metrics, AE-F1, AS-F1, and Overall F1 in evaluation. Three strong baseline models in the main results, namely SPAN, IMN, and RACL, are compared here. As the results demonstrated in Table 3, our RF-MRC achieves the best performance. This suggests that RF-MRC still works in more detailed and complex sentences. It is interesting to observe that AS-F1 improves more than AE-F1 in this comparison. We conjecture this is because our model can capture relations between aspect terms and potential opinion terms, even if there are no opinion annotations in MAMS.

4.6 Ablation Test

In order to investigate the effect of the query answer flipped process, we perform comprehensive ablation studies on three datasets. Table 4 shows the results of the Overall-F1 measure. We remove



(a) The effect of p .

(b) The effect of T .

Figure 4: Effects of parameters.

the process ‘‘A2O’’ and ‘‘O2A’’, respectively, and derive two degraded variants denoted by ‘‘w/o A2O’’ and ‘‘w/o O2A’’. As expected, both of ‘‘A2O’’ and ‘‘O2A’’ processes are effective for the whole task. It is noted that scores of the model without ‘‘A2O’’ decrease more than those of the model without ‘‘O2A’’ on Restaurant14 and Laptop14. We consider it is probably because the extraction of ATE is more accurate than OTE on the two datasets, which can be discovered in Table 2. The model ‘‘w/o A2O’’ performs better than ‘‘w/o O2A’’ since the OTE on Restaurant15 is more accurate than ATE (c.f. Table 2).

4.7 Effect of Parameters

Next, we study the effects of different hyperparameters in our model, including the number of the candidate aspect terms p , and the round of cross question answering T , to evaluate how they contribute to the performance. We exhibit the overall F1 in Figure 4. Because the effect of q , which is the number of candidate opinion terms, is similar to p , we omit the repeated display.

As Figure 4(a) shows, the model performs best on Restaurant14 and Laptop14 when $p = 8$. We believe that the model ignores some true aspects

Case	IMN	RACL	RF-MRC
The (outdoor patio) _{pos} is really nice in good weather, but what (ambiance) _{neu} the indoors possesses is negated by the noise and the crowds.	(outdoor patio) _{pos} (crowds) _{neg} ✗	(outdoor patio) _{pos} (null) ✗	(outdoor patio) _{pos} (ambiance) _{neu}
The (food) _{pos} is pretty good, but after 2 or 3 bad experiences at the restaurant (consistently rude, late with RSVP'd (seating) _{neu}), I decided I would only order (delivery) _{neu} .	(food) _{pos} (seating) _{neg} (null) ✗	(food) _{pos} (seating) _{neg} ✗ (delivery) _{neu}	(food) _{pos} (seating) _{neu} (delivery) _{neu}
(Dinner) _{neu} is okay not many vegetarian options and the (portions) _{neg} are small.	(Dinner) _{pos} ✗ (portions) _{neg}	(Dinner) _{pos} ✗ (vegetarian options) _{neg} ✗ (portions) _{neg}	(Dinner) _{pos} ✗ (vegetarian options) _{neg} ✗ (portions) _{neg}

Table 5: Case Study. The abbreviations *pos*, *neu* and *neg* on the table represent positive, neutral and negative sentiments, respectively. The sentiment polarities are demonstrated as the subscripts of aspect terms. “null” denotes that there is an aspect which is not extracted.

when p decreases, while more inaccurate aspects will be taken into consideration with the value of p increasing. In Figure 4(b), the model is less effective when $t = 1$ while the performance is best when $t = 2$. When $t > 2$, the Overall-F1 shows a decreasing trend. It is possible that too many rounds of question answering are prone to overfitting.

4.8 Case Study

Finally, we conduct a case study to illustrate the effectiveness and perform an error analysis. We select three cases from the MAMS dataset and compare our results with IMN and RACL. Table 5 reports the results.

In the first case, there are two aspects, i.e., “outdoor patio” and “ambiance”. Both IMN and RACL cannot identify “ambiance” as aspect terms. We conjecture the possible reasons might be they only consider relations between sentence representations of sub-tasks, which derives the aspect term “ambiance” is weakened in such a complex sentence. In addition, IMN extracts “crowds” as an extra aspect might because it fails to consider the relations between aspect terms and relevant opinion terms. However, our proposed model extracts all the aspect terms and predicts corresponding sentiment polarities correctly.

The second case is a longer sentence with three aspects and expresses positive and neutral polarities. Our RF-MRC extracts all aspect terms and opinion terms and predicts corresponding polarities successfully. However, IMN can not extract “delivery” and we conjecture the performance on ATE decreases in a longer sentence. RACL extracts all aspect terms correctly but the polarity of “seating” is misjudged. Because RACL exploits different

semantic relations between sub-tasks, it is possible that it captures the inaccurate “rude” and “late” as evidence to predict the sentiment for “seating” as “negative”. This case demonstrates that the proposed model has more advantages to solve complex sentences.

We perform an error analysis in the third case. We see that the demonstrated sentence is much shorter than the former two. However, all the three models predict the wrong sentiment for the aspect “dinner”. We analyze it is because the “okay” is regarded as the opinion word for “dinner”, and this word may usually represent positive polarity in the training set. Recall that our training loss of cross-entropy seeking for a maximum likelihood in the training set, which might be that the reason for deriving a wrong prediction in this case. More interestingly, RACL and our RF-MRC, as two SOTA solutions, extract “vegetarian options” as an aspect incorrectly. By looking closer at this sentence, we find that the seldom choice in “vegetarian options” is evidence of why the user says “dinner” is just okay. Hence, understanding the structure of sentences by logical even causal inference might be shed new light on future research of this area.

Moreover, we select a sentence from the test set of Restaurant14 and present visualization of the extraction results and the matching process in Figure 5, successively. Specifically, the aspect terms are marked as red while opinion terms are marked in blue. According to Figure 5(a) and 5(b), we can see our RF-MRC can accurately extract aspect terms, i.e., “food” and “waiting”, and opinion terms, i.e., “good”, “popular” and “nightmare”. As Figure 5(c) shown, the “food” has higher scores with “good” and “popular” while the “waiting” is more relevant to “nightmare”. Based on the obser-

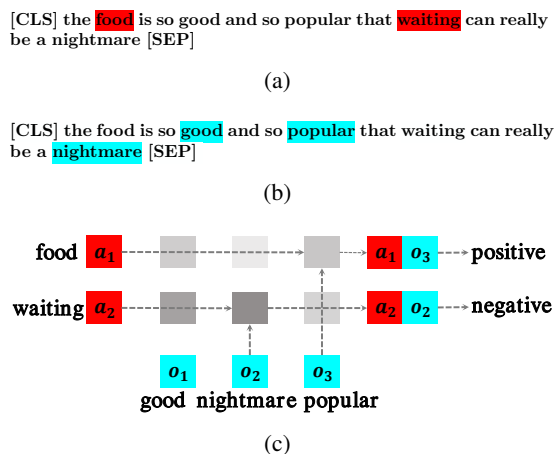


Figure 5: An example of extraction results and matching process.

variations, we can infer that the proposed RF-MRC is capable of associating the aspect terms with relevant opinion terms and matching them in pairs for sentiment classification.

5 Conclusion

In this paper, we investigate the unified ABSA from the perspective of MRC and propose a new paradigm named RF-MRC. Either extracted aspect terms or opinion terms are constructed as queries, and the related opinion terms and aspect terms are considered as answers. We further design a matching module to match all the extracted aspect terms and relevant opinion terms, and predict the sentiment polarities. Experiments on three widely used benchmarks and a challenging dataset demonstrate the superiority of the proposed framework.

Acknowledgements

This work is supported by the National Key Research and Development Program of China under Grant No. 2017YFB1002104, the National Natural Science Foundation of China under Grant No. 92046003, 61976204, U1811461. Xiang Ao is also supported by the Project of Youth Innovation Promotion Association CAS and Beijing Nova Program Z201100006820062. We would like to thank the anonymous reviewers for their valuable comments, and Mengda Huang, Hai Jia for their insightful discussions.

References

Stefanos Angelidis and Mirella Lapata. 2018. [Summarizing opinions: Aspect extraction meets senti-](#)

[ment prediction and they are both weakly supervised.](#) In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3675–3686, Brussels, Belgium. Association for Computational Linguistics.

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.

Zhuang Chen and Tiejun Qian. 2019. [Transfer capsule network for aspect level sentiment classification.](#) In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 547–556, Florence, Italy. Association for Computational Linguistics.

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.

Hongliang Dai and Yangqiu Song. 2019. [Neural aspect and opinion term extraction with mined rules as weak supervision.](#) In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5268–5277, Florence, Italy. 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.

Chunning Du, Haifeng Sun, Jingyu Wang, Qi Qi, Jianxin Liao, Tong Xu, and Ming Liu. 2019. [Capsule network with interactive attention for aspect-level sentiment classification.](#) In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5489–5498, Hong Kong, China. Association for Computational Linguistics.

Luheng He, Mike Lewis, and Luke Zettlemoyer. 2015. [Question-answer driven semantic role labeling: Using natural language to annotate natural language.](#) In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 643–653, Lisbon, Portugal. Association for Computational Linguistics.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2018. [Effective attention modeling for](#)

- aspect-level sentiment classification. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1121–1131, Santa Fe, New Mexico, USA. Association for Computational Linguistics.
- 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 *ACL*, pages 504–515.
- Xiaochen Hou, Jing Huang, Guangtao Wang, Peng Qi, Xiaodong He, and Bowen Zhou. 2021a. [Selective attention based graph convolutional networks for aspect-level sentiment classification](#). In *Proceedings of the Fifteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-15)*, pages 83–93, Mexico City, Mexico. Association for Computational Linguistics.
- Xiaochen Hou, Peng Qi, Guangtao Wang, Rex Ying, Jing Huang, Xiaodong He, and Bowen Zhou. 2021b. [Graph ensemble learning over multiple dependency trees for aspect-level sentiment classification](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2884–2894, Online. 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.
- Binxuan Huang and Kathleen Carley. 2019. [Syntax-aware aspect level sentiment classification with graph attention networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5469–5477, Hong Kong, China. Association for Computational Linguistics.
- Jiaxin Huang, Yu Meng, Fang Guo, Heng Ji, and Jiawei Han. 2020. [Weakly-supervised aspect-based sentiment analysis via joint aspect-sentiment topic embedding](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6989–6999, Online. Association for Computational Linguistics.
- Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. [A challenge dataset and effective models for aspect-based sentiment analysis](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6280–6285, Hong Kong, China. Association for Computational Linguistics.
- Diederik P Kingma and Jimmy Ba. 2015. [Adam: A method for stochastic optimization](#). In *International Conference on Learning Representations*.
- Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. [Zero-shot relation extraction via reading comprehension](#). In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342, Vancouver, Canada. Association for Computational Linguistics.
- Kun Li, Chengbo Chen, Xiaojun Quan, Qing Ling, and Yan Song. 2020a. [Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7056–7066, Online. Association for Computational Linguistics.
- Xiaoya Li, Jingrong Feng, Yuxian Meng, Qinghong Han, Fei Wu, and Jiwei Li. 2020b. [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.
- Xiaoya Li, Fan Yin, Zijun Sun, Xiayu Li, Arianna Yuan, Duo Chai, Mingxin Zhou, and Jiwei Li. 2019a. [Entity-relation extraction as multi-turn question answering](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1340–1350, Florence, Italy. Association for Computational Linguistics.
- Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. [Transformation networks for target-oriented sentiment classification](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 946–956, Melbourne, Australia. Association for Computational Linguistics.
- Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019b. [A unified model for opinion target extraction and target sentiment prediction](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 6714–6721.
- 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.
- Bin Liang, Jiachen Du, Ruifeng Xu, Binyang Li, and Hejiao Huang. 2019. [Context-aware embedding for targeted aspect-based sentiment analysis](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4678–4683, Florence, Italy. Association for Computational Linguistics.

- Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167.
- Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019a. **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.
- Ling Luo, Xiang Ao, Yan Song, Jinyao Li, Xiaopeng Yang, Qing He, and Dong Yu. 2019b. **Unsupervised neural aspect extraction with sememes**. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 5123–5129. International Joint Conferences on Artificial Intelligence Organization.
- 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, 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**. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 13543–13551.
- Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2018. **The natural language decathlon: Multitask learning as question answering**. *arXiv preprint arXiv:1806.08730*.
- 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.
- Minh Hieu Phan and Philip O. Ogunbona. 2020. **Modelling context and syntactical features for aspect-based sentiment analysis**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3211–3220, Online. Association for Computational Linguistics.
- Maria Pontiki, Dimitrios Galanis, Harris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. **Semeval-2015 task 12: Aspect based sentiment analysis**. In *SemEval 2015*, pages 486–495.
- Maria Pontiki, Haris Papageorgiou, Dimitrios Galanis, Ion Androutsopoulos, John Pavlopoulos, and Suresh Manandhar. 2014. **Semeval-2014 task 4: Aspect based sentiment analysis**. *SemEval 2014*, page 27.
- Pranav Rajpurkar, Robin Jia, and Percy Liang. 2018. **Know what you don’t know: Unanswerable questions for SQuAD**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 784–789, Melbourne, Australia. Association for Computational Linguistics.
- Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. **SQuAD: 100,000+ questions for machine comprehension of text**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392, Austin, Texas. Association for Computational Linguistics.
- Kai Sun, Richong Zhang, Samuel Mensah, Yongyi Mao, and Xudong Liu. 2019. **Aspect-level sentiment analysis via convolution over dependency tree**. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5679–5688, Hong Kong, China. Association for Computational Linguistics.
- Hao Tang, Donghong Ji, Chenliang Li, and Qiji Zhou. 2020. **Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6578–6588, Online. Association for Computational Linguistics.
- Jialong Tang, Ziyao Lu, Jinsong Su, Yubin Ge, Linfeng Song, Le Sun, and Jiebo Luo. 2019. **Progressive self-supervised attention learning for aspect-level sentiment analysis**. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 557–566, Florence, Italy. Association for Computational Linguistics.
- Hao Tian, Can Gao, Xinyan Xiao, Hao Liu, Bolei He, Hua Wu, Haifeng Wang, and Feng Wu. 2020. **SKEP: Sentiment knowledge enhanced pre-training for sentiment analysis**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4067–4076, Online. Association for Computational Linguistics.
- Feixiang Wang, Man Lan, and Wenting Wang. 2018. **Towards a one-stop solution to both aspect extraction and sentiment analysis tasks with neural multi-task learning**. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE.
- Kai Wang, Weizhou Shen, Yunyi Yang, Xiaojun Quan, and Rui Wang. 2020. **Relational graph attention network for aspect-based sentiment analysis**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3229–3238, Online. Association for Computational Linguistics.
- Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. **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 AAAI Conference on Artificial Intelligence*, volume 31.
- Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016b. [Attention-based LSTM for aspect-level sentiment classification](#). In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.
- Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, and Jiwei Li. 2020. [CorefQA: Coreference resolution as query-based span prediction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6953–6963, Online. Association for Computational Linguistics.
- Chi Xu, Hao Feng, Guoxin Yu, Min Yang, Xiting Wang, Yan Song, and Xiang Ao. 2021. [Discovering protagonist of sentiment with aspect reconstructed capsule network](#). In *International Conference on Database Systems for Advanced Applications*, pages 120–135. Springer.
- 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.
- Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. [Position-aware tagging for aspect sentiment triplet extraction](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2339–2349, Online. Association for Computational Linguistics.
- Wei Xue and Tao Li. 2018. [Aspect based sentiment analysis with gated convolutional networks](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2514–2523, Melbourne, Australia. Association for Computational Linguistics.
- Guoxin Yu, Xiang Ao, Ling Luo, Min Yang, Xiaofei Sun, Jiwei Li, and Qing He. 2021. [Making flexible use of subtasks: A multiplex interaction network for unified aspect-based sentiment analysis](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2695–2705, Online. Association for Computational Linguistics.
- Chen Zhang, Qiuchi Li, and Dawei Song. 2019. [Aspect-based sentiment classification with aspect-specific graph convolutional networks](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4568–4578, Hong Kong, China. Association for Computational Linguistics.
- Meishan Zhang, Yue Zhang, and Duy-Tin Vo. 2016. [Gated neural networks for targeted sentiment analysis](#). In *Thirtieth AAAI conference on artificial intelligence*.
- 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, Online. Association for Computational Linguistics.