COM-MRC: A COntext-Masked Machine Reading Comprehension Framework for Aspect Sentiment Triplet Extraction

Zepeng Zhai[1], **Hao Chen**[1], **Fangxiang Feng**[1,2], **Ruifan Li**[1,2]*, **Xiaojie Wang**[1,2]
[1]School of Artificial Intelligence, Beijing University of Posts and Telecommunications, China
[2]Engineering Research Center of Information Networks, Ministry of Education, China
`{zepeng, ccchenhao997, fxfeng, rfli, xjwang}@bupt.edu.cn`

Abstract

Aspect Sentiment Triplet Extraction (ASTE) aims to extract sentiment triplets from sentences, which was recently formalized as an effective machine reading comprehension (MRC) based framework. However, when facing multiple aspect terms, the MRC-based methods could fail due to the interference from other aspect terms. In this paper, we propose a novel *COntext-Masked MRC* (COM-MRC) framework for ASTE. Our COM-MRC framework comprises three closely-related components: a context augmentation strategy, a discriminative model, and an inference method. Specifically, a context augmentation strategy is designed by enumerating all masked contexts for each aspect term. The discriminative model comprises four modules, i.e., aspect and opinion extraction modules, sentiment classification and aspect detection modules. In addition, a two-stage inference method first extracts all aspects and then identifies their opinions and sentiment through iteratively masking the aspects. Extensive experimental results on benchmark datasets show the effectiveness of our proposed COM-MRC framework, which outperforms state-of-the-art methods consistently[1].

1 Introduction

Aspect Sentiment Triplet Extraction (ASTE) has recently been proposed, which is a variant of the fine-grained Aspect-based Sentiment Analysis (ABSA) task. For a given sentence, ASTE aims to extract the sentiment triplets, including aspect term, opinion term and the corresponding sentiment polarity. As shown in Figure 1, ASTE could produce two triplets from the given sentence.

For ASTE task, early methods adopt a two-stage pipeline framework that first identifies aspects with sentiment and opinions then pairs them, producing

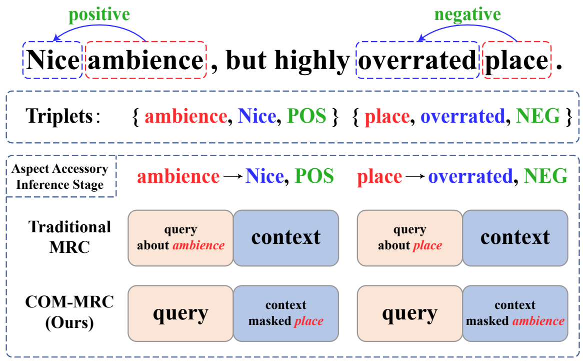


Figure 1: A sentence and its sentiment triplets are shown at the top half. Moreover, the primary difference between the traditional MRC and our COM-MRC is highlighted at the bottom.

the sentiment triplets (Peng et al., 2020). However, these pipeline-based methods ignore the interaction among triplets, which could result in the error propagation. To alleviate this problem, some recent studies jointly extract the sentiment triplets in an end-to-end framework (Xu et al., 2020; Wu et al., 2020a; Zhang et al., 2020; Chen et al., 2021b; Yan et al., 2021), which is constructed mainly by designing a tagging scheme.

Very recently, Mao et al. (2021) and Chen et al. (2021a) formalized ASTE by using a machine reading comprehension (MRC) framework. The basic idea of MRC-based methods is using multi-turn QA under an identical context with diverse queries. Specifically, MRC-based methods involve two stages, Aspect Inference **(AI)** and Aspect Accessory Inference **(AAI)**. The former is to extract aspect terms by constructing a query about aspects, e.g., "*What aspects?*". The latter is to identify the corresponding opinions and sentiment by constructing queries for each aspect term, e.g., "*What opinions and sentiment given the aspect ambience?*".

Despite the impressive performance, however, MRC-based methods may suffer the interference problem when analyzing sentences with multiple aspects. Intuitively, the more aspects a sentence

*Corresponding author.
[1]Code and datasets are available at `https://github.com/zzp-seeker/COM-MRC`.

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, pages 3230–3241
December 7-11, 2022 ©2022 Association for Computational Linguistics

Multi-aspect	Rest 14	Lap 14	Rest 15	Rest 16
Sentence	42.7%	29.4%	29.2%	28.7%
Triplet	62.9%	48.1%	47.2%	46.6%

Table 1: The proportion of sentences containing multiple aspect terms, and that of triplets in these sentences for the benchmark dataset proposed by Xu et al. (2020).

contains, the harder it is usually to obtain the correct correspondence between aspects and their accessories. For the example in Figure 1, the model may mistakenly match the opinion "*overrated*" to "*ambience*". Moreover, the trained model equipped with attention mechanism would usually capture the potential correlation between aspects and opinions. If the model pays attention to other aspects, it will also attend to the opinions of these aspects, which will interfere with the opinion inference process of the current aspect. As shown in the example in Figure 1, for the aspect "*ambience*", the model may pay more attention to the incorrect opinion "*overrated*" if the aspect "*place*" can be attended. Note that statistics in Table 1 on benchmark datasets show that remark sentences with multiple aspect terms occupy a large portion. The sentiment triplets in these sentences account for roughly a half. Therefore, how to identify the information from different aspects more effectively and further alleviate the interference of unrelated aspects is challenging. Motivated by above observations, we present the idea of masking aspects for alleviating the interference problem.

In this paper, we propose a novel framework called COntext-Masked machine reading comprehension (COM-MRC) for ASTE. Our COM-MRC framework is in general based on the idea of masking aspect, and it comprises three closely-related components: a context augmentation strategy, a discriminative model, an inference method. **Firstly**, to alleviate the interference and to better identify information from different aspects, we use the idea of masking aspects for context augmentation. We argue that a sentence with multiple aspect terms deserves to be treated as multiple training samples due to the difficulty of extracting triplets. Hence, we set a regular query with various masked contexts to identify each aspect and its accessories. For a sentence with t aspect terms, the number of samples grows from 1 to 2^t. Thus, the training corpus effectively expands. **Secondly**, to effectively capture the correlation among sentiment triplets, we design a discriminative model. An aspect detection module detects whether there exist aspect terms in the masked context. This module and the other three modules, i.e., aspect extraction, opinion extraction, and sentiment classification modules, work collaboratively for ASTE task. **Thirdly**, the aspect is extracted one by one from left to right during AI stage through iteratively masking aspects. Then, during AAI stage, all other unrelated aspects are masked in the context for more precise identification. The three components constitute our COM-MRC framework which could alleviate the interference problem. Experimental results show that our COM-MRC framework consistently outperforms state-of-the-art methods.

Our contributions are summarized as follows.

- We propose a novel COM-MRC framework for ASTE task. Our framework comprises three components: a context augmentation strategy, a discriminative model, and an inference method.
- We use the context augmentation strategy to obtain effective expansion of the training corpus. We design the discriminative model with four collaborative modules. We implement our inference method by iteratively masking aspects.
- We conduct extensive experiments on two groups of benchmark datasets. The experimental results demonstrate the effectiveness of our COM-MRC framework. The source code and data of our work are released for knowledge sharing.

2 Proposed COM-MRC Framework

2.1 Problem Formulation

Given a sentence $S = \{w_1, w_2, ..., w_n\}$ with n tokens, the aim of ASTE task is to extract all sentiment triplets $\mathcal{T}$ within the sentence. Each sentiment triplet is represented as a tuple (a, o, s), where symbols a, o, and s represent the aspect term, the opinion term and the sentiment polarity, respectively. The range of sentiment polarity is given as three types, i.e., $s \in \{POS, NEU, NEG\}$.

2.2 Context Augmentation Strategy

Primarily, our discriminative model takes a fixed query and a masked context as input. We then adopt BERT (Devlin et al., 2019) as the sentence encoder to represent their semantics.

Specifically, we devise a fixed query to prompt our model for adapting ASTE task. Here, we identify the leftmost aspect and its corresponding opin-

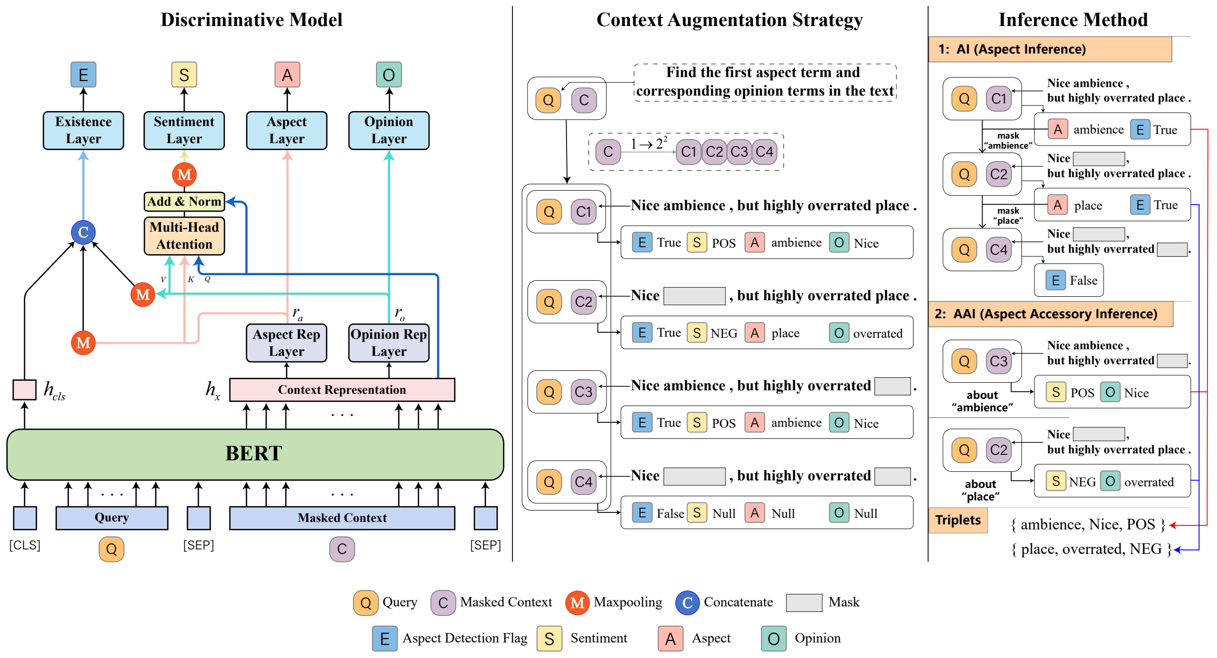


Figure 2: The overview of our COM-MRC framework. The discriminative model is given on the left. The context augmentation strategy is illustrated in the middle. The inference method involving two stages is depicted.

ion terms. The query q is given as follows:

$$q = \text{“}\textit{Find the first aspect term and corresponding opinion terms in the text}\text{”} \quad (1)$$

Strategy. For the contexts, we design an augmentation strategy. Suppose that a sentence S consists of t aspect terms. For each aspect term, we perform two types of operations, i.e., *masking* or *not masking*. Thus, one training sentence expands to 2^t instances. This augmentation strategy is illustrated in Figure 2.

Specifically, we mask the k-th token by setting its attention score to 0. A masking matrix M is accordingly defined as follows,

$$M_{ij} = \begin{cases} -\infty, & \text{if } j = k \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

Then, we apply the matrix to the attention module A in BERT given the query Q, the key K and the value V as follows,

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}} + M\right) V \quad (3)$$

where d is the dimension of the key.

With the fixed query q in Eq. (1) and a masked context x produced using the aforementioned strategy, we then adopt BERT (Devlin et al., 2019) to represent their semantics. The specific input is given as *“[CLS] q [SEP] x [SEP]”*. Suppose that the query q contains m tokens. Note that the masked context x contains the same length of tokens n as that of the sentence. We obtain the representation $h \in \mathbb{R}^{d \times (m+n+3)}$ from the last BERT block. The representations of the masked context and the token *[CLS]* are denoted as $h_x \in \mathbb{R}^{d \times n}$ and $h_{cls} \in \mathbb{R}^{d \times 1}$, respectively.

2.3 Discriminative Model

Our discriminative model comprises four modules. The structure is depicted in Figure 2.

Aspect Extraction Module. To obtain the first unmasked aspect term, motivated by span-based methods (Hu et al., 2019), we obtain the probabilities for starting and ending positions from the context representation h_x as follows:

$$r_a = W_{a,1} h_x \quad (4)$$
$$p^{a,s} = \text{softmax}(W_{a,2} r_a) \quad (5)$$
$$p^{a,e} = \text{softmax}(W_{a,3} r_a) \quad (6)$$

where $W_{a,1} \in \mathbb{R}^{d \times d}$, $W_{a,2} \in \mathbb{R}^{1 \times d}$, and $W_{a,3} \in \mathbb{R}^{1 \times d}$ are trainable parameters. In addition, $r_a \in \mathbb{R}^{d \times n}$ stands for the representation of aspect term. Correspondingly, we use the cross-entropy as the loss function for starting and ending positions. The aspect extraction loss $\mathcal{L}_A$ is defined as follows:

$$\mathcal{L}_A = -\sum_{i=1}^{n} y_i^{a,s} \log p_i^{a,s} - \sum_{i=1}^{n} y_i^{a,e} \log p_i^{a,e} \quad (7)$$

where $y^{a,s}$ and $y^{a,e} \in \mathbb{R}^n$ are ground truths of starting and ending positions for the first unmasked aspect term. The subscript i denotes the i-th token.

Opinion Extraction Module. To obtain all opinion terms for the first unmasked aspect, we build an opinion extraction module similar to aspect extraction module. Thus, we obtain the opinion representation r_o and the module's loss function $\mathcal{L}_O$.

Sentiment Classification Module. Intuitively, the sentiment polarity is highly related to the masked context, the aspect term, and the opinion term. In our model, we use multi-head attention (Vaswani et al., 2017) to fuse these three semantic information. This process is formulated as follows:

$$r_s = \mathrm{LN}(h_x + \mathrm{MultiHead}(h_x, r_a, r_o)) \quad (8)$$

$$g_s = \mathrm{MP}(r_s) \quad (9)$$

$$p^s = \mathrm{softmax}(W_s g_s + b_s) \quad (10)$$

where LN, MultiHead, and MP represent three operations, i.e., layer norm, multi-head attention and max pooling, respectively. In addition, $r_s \in \mathbb{R}^{d \times n}$ and $g_s \in \mathbb{R}^{d \times 1}$ are intermediate variables. $W_s \in \mathbb{R}^{3 \times d}$ and b_s are the trainable weight and bias, respectively. The cross entropy loss $\mathcal{L}_S$ for the sentiment classification is then given as follows:

$$\mathcal{L}_S = -\sum_{i=1}^{3} y_i^s \log p_i^s \quad (11)$$

where $y^s \in \mathbb{R}^3$ is the label of sentiment polarity.

Aspect Detection Module. This module is to detect whether there exist aspect terms in the masked context. For the context with all aspect terms being masked, its label is set *False*, otherwise *True*. The module works according to the *[CLS]* token representation h_{cls}, aspect representation r_a, and opinion representation r_o as follows:

$$r_e = h_{cls} \oplus \mathrm{MP}(r_o) \oplus \mathrm{MP}(r_a) \quad (12)$$

$$p^e = \mathrm{softmax}(W_e r_e + b_e) \quad (13)$$

where $r_e \in \mathbb{R}^{3d \times 1}$ is an intermediate variable. In addition, $W_e \in \mathbb{R}^{2 \times 3d}$ and b_e are the trainable weight and bias, respectively. The symbol $\oplus$ means the concatenation operation. We use the binary cross entropy loss $\mathcal{L}_E$ for aspect detection module.

Algorithm 1 Inference Algorithm

Input: Sentence S and query q.
Output: Triplets $\mathcal{T} = \{(a, o, s)\}_N$.
1: Initialize $\mathcal{T}, \mathcal{A} = \{\}, \{\}$;
   *// **AI Stage**: Get aspect detection flag e and aspect a*
2: $e, a \leftarrow \mathrm{GetAI}(q, S)$
3: **while** $e = \mathrm{True}$ **do**
4:     $\mathcal{A} \leftarrow \mathcal{A} \cup \{a\}$
5:     $e, a \leftarrow \mathrm{GetAI}(q, S.\mathrm{Mask}(\mathcal{A}))$
6: **end while**
   *// **AAI Stage**: Get opinion set O and sentiment s*
7: **for** $a_i \in \mathcal{A}$ **do**
8:     $O, s \leftarrow \mathrm{GetAAI}(q, S.\mathrm{Mask}(\mathcal{A} - \{a_i\}))$
9:     **for** $o_j \in O$ **do**
10:         $\mathcal{T} \leftarrow \mathcal{T} \cup \{(a_i, o_j, s)\}$
11:     **end for**
12: **end for**
13: **return** $\mathcal{T}$

Loss Function. At last, our objective is to minimize the following total loss:

$$\mathcal{L}_T = \alpha\mathcal{L}_A + \beta\mathcal{L}_O + \gamma\mathcal{L}_S + \delta\mathcal{L}_E \quad (14)$$

where α, β, γ and δ are four hyper-parameters used to adjust the influence of the corresponding losses.

2.4 Inference Method

To alleviate the interference from all the other aspects, we present our inference method. Our method involves two successive stages, AI and AAI. AI stage is to extract all aspects; AAI stage is to identify the opinions and sentiment polarities for all the aspects. In Figure 2, we give an example to illustrate the two stages.

During **AI** stage, firstly, we obtain the aspect detection flag e and the first aspect term a using our trained model for the query q and the sentence S. If the detection flag e is True, we append the aspect a to aspect set $\mathcal{A}$ and then mask aspect a in the sentence S. With the query and the masked context, we again use the trained model to obtain the next aspect detection flag and the next aspect. We repeat the above step until the detection flag e is False. At last, we obtain the aspect set $\mathcal{A}$.

During **AAI** stage, to produce the context for one aspect term a, we mask all the aspects except a. Combined with the fixed query q, the masked context is fed into our trained model. Thus, for that aspect term we obtain its opinion term set O and the corresponding sentiment polarity s. Finally, based on the set O, we append all the triplets to the triplet set $\mathcal{T}$. The inference method is formally summarized in Algorithm 1. Here, $S.\mathrm{Mask}(\mathcal{A})$ in Line 5 means that the sentence S is updated to masked context by masking all the aspects belonging to the current aspect set $\mathcal{A}$.

Dataset		Rest 14				Lap 14				Rest 15				Rest 16			
		#S	#MA-S	#T	#MA-T	#S	#MA-S	#T	#MA-T	#S	#MA-S	#T	#MA-T	#S	#MA-S	#T	#MA-T
$\mathcal{D}_1$	train	1259	536	2356	1485	899	254	1452	685	603	190	1038	512	863	251	1421	669
	dev	315	119	580	322	225	75	383	208	151	42	239	108	216	62	348	162
	test	493	228	1008	668	332	103	547	266	325	82	493	213	328	93	525	238
$\mathcal{D}_2$	train	1266	533	2338	1443	906	265	1460	709	605	183	1013	489	857	244	1394	652
	dev	310	123	577	352	219	59	346	155	148	49	249	125	210	65	339	163
	test	492	228	994	662	328	103	543	266	322	82	485	211	326	91	514	232

Table 2: Statistics for the two groups of experimental datasets, $\mathcal{D}_1$ and $\mathcal{D}_2$. #S and #T denote the number of sentences and triplets, respectively. #MA-S denotes the number of sentences containing multiple aspect terms. #MA-T denotes the number of triplets in the corresponding sentences containing multiple aspects.

3 Experiments

3.1 Datasets

We conduct experiments on two groups of benchmark datasets for ASTE. These datasets were created from the SemEval Challenges (Pontiki et al., 2014, 2015, 2016). The first group $\mathcal{D}_1$[2] including four subsets (Rest 14, Lap 14, Rest 15, and Rest 16) is annotated by Wu et al. (2020a). The second group $\mathcal{D}_2$[3] is proposed by Xu et al. (2020), which is a corrected version of dataset annotated by Peng et al. (2020). Table 2 shows the statistics of these two groups of datasets.

3.2 Baseline Methods

We compare our COM-MRC with state-of-the-art baselines. These baseline models are briefly categorized into the following three groups. 1) **Pipeline.** CMLA+, RINANTE+, Li-unified-R, and Peng-two-stage are proposed by Peng et al. (2020). Peng-two-stage+IOG and IMN+IOG are proposed by Wu et al. (2020a). 2) **End-to-end.** This group includes OTE-MTL (Zhang et al., 2020), JET-BERT (Xu et al., 2020), GTS (Wu et al., 2020a), S^3E^2 (Chen et al., 2021b), Unified (Yan et al., 2021), SPAN-ASTE (Xu et al., 2021) and EMC-GCN (Chen et al., 2022). 3) **MRC-based.** BMRC (Chen et al., 2021a) devises three types of queries to build the associations among different subtasks based on MRC.

3.3 Implementation Details

We use the Bert-Base-Uncased English version[4] as our base encoder. Our model is trained for 100 epochs with a linear warmup for 10% of training steps followed by a cosine decay of learning rate to 0. AdamW optimizer (Loshchilov and Hutter, 2019) is used with the maximum learning rate of 9×10^{-5} for BERT weights and weight decay of 10^{-2}. The batch size is 15, and the dropout rate is set to 0.1. Considering the prediction performance with masked contexts will be greatly affected if the detected aspect terms are incorrect, we set a larger weight for aspect extraction in the loss function for more accurate identification of aspect term. Four hyper-parameters α, β, γ and δ in Eq. (14) are set to 8.0, 3.2, 1.0 and 1.0 respectively. We use a heuristic multi-span decoding algorithm (Hu et al., 2019) to obtain the aspect and opinion spans during inference and the threshold is manually set. We use a GeForce RTX 3090 to train the model for an average of 0.85h. We save the model parameters according to the model's best performance on the development set. The reported results are the averages on five runs with different random seeds.

3.4 Main Results

We compare our COM-MRC with other baselines in terms of Precision, Recall and F1 scores. The experimental results on $\mathcal{D}_1$ and $\mathcal{D}_2$ are reported in Tables 3 and 4, respectively. Under F1 metric, our COM-MRC consistently outperforms all pipeline, end-to-end and MRC-based methods on the two groups of datasets. Note that our method outperforms the best end-to-end method SPAN-ASTE on $\mathcal{D}_2$. We observe that the end-to-end and MRC-based methods are more competitive than the pipeline methods as they alleviate the error propagation and establish the correlations between related subtasks. Moreover, compared with another strong MRC-based method, i.e., BMRC, our COM-MRC significantly surpasses its performance by an average of 3.59% and 4.18% F1-score on $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. This improvement is attributed to that our COM-MRC can effectively alleviate the interference problem via a context augmentation strategy, a discriminative model, and an inference method. In addition, in order to show the signifi-

[2]https://github.com/NJUNLP/GTS

[3]https://github.com/xuuuluuu/SemEval-Triplet-data/tree/master/ASTE-Data-V2-EMNLP2020

[4]https://github.com/huggingface/transformers

Model	Rest 14			Lap 14			Rest 15			Rest 16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Li-unified-R	41.44	68.79	51.68	42.25	42.78	42.47	43.34	50.73	46.69	38.19	53.47	44.51
Peng-two-stage	44.18	62.99	51.89	40.40	47.24	43.50	40.97	54.68	46.79	46.76	62.97	53.62
Peng-two-stage+IOG	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.04	48.71	59.25	58.09	58.67
IMN+IOG	59.57	63.88	61.65	49.21	46.23	47.68	55.24	52.33	53.75	-	-	-
S^3E^2	69.08	64.55	66.74	59.43	46.23	52.01	61.06	56.44	58.66	71.08	63.13	66.87
GTS-BiLSTM	67.28	61.91	64.49	59.42	45.13	51.30	63.26	50.71	56.29	66.07	65.05	65.56
GTS-CNN	70.79	61.71	65.94	55.93	47.52	51.38	60.09	53.57	56.64	62.63	66.98	64.73
GTS-BERT	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	68.58	66.60	67.58
BMRC	-	-	70.01	-	-	57.83	-	-	58.74	-	-	67.49
EMC-GCN	71.85	72.12	71.98	61.46	55.56	58.32	59.89	61.05	60.38	65.08	71.66	68.18
Our COM-MRC	76.45	69.67	**72.89**	64.73	56.09	**60.09**	68.50	59.74	**63.65**	72.80	70.85	**71.79**

Table 3: Results on the benchmark $\mathcal{D}_1$ (Wu et al., 2020a). All baseline results are copied from the original papers.

Model	Rest 14			Lap 14			Rest 15			Rest 16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+†	39.18	47.13	42.79	30.09	36.92	33.16	34.56	39.84	37.01	41.34	42.10	41.72
RINANTE+†	31.42	39.38	34.95	21.71	18.66	20.07	29.88	30.06	29.97	25.68	22.30	23.87
Li-unified-R†	41.04	67.35	51.00	40.56	44.28	42.34	44.72	51.39	47.82	37.33	54.51	44.31
Peng-two-stage†	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	46.96	64.24	54.21
OTE-MTL*	62.00	55.97	58.71	49.53	39.22	43.42	56.37	40.94	47.13	62.88	52.10	59.96
JET-BERT†	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
GTS-BERT*	68.09	69.54	68.81	59.40	51.94	55.42	59.28	57.93	58.60	68.32	66.86	67.58
Unified	65.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
BMRC*	75.61	61.77	67.99	70.55	48.98	57.82	68.51	53.40	60.02	71.20	61.08	65.75
SPAN-ASTE	72.89	70.89	71.85	63.44	55.84	59.38	62.18	64.45	63.27	69.45	71.17	70.26
EMC-GCN	71.21	72.39	71.78	61.70	56.26	58.81	61.54	62.47	61.93	65.62	71.30	68.33
Our COM-MRC	75.46	68.91	**72.01**	62.35	58.16	**60.17**	68.35	61.24	**64.53**	71.55	71.59	**71.57**

Table 4: Results on the benchmark $\mathcal{D}_2$ (Xu et al., 2020). The symbol † means that the results are retrieved from Xu et al. (2020). The symbol * denotes that the corresponding results are retrieved from Chen et al. (2022).

cance of our experimental results, we conduct pairwise t-test on F1 comparing our COM-MRC with BMRC and EMC-GCN on two datasets, $\mathcal{D}_1$ and $\mathcal{D}_2$. All of the produced p-values are less than 0.05.

4 Analysis

4.1 On Context Augmentation Strategy

Is the strategy on context augmentation effective? We conduct experiments compared with another two strategies. The linear strategy is only considering contexts to be used in the inference process. For a sentence with t aspects, this method produces $2t$ samples[5]. The NOP strategy is using the original sentences without augmentation. For the exponential strategy used in our COM-MRC, we obtain 2^t samples for one sentence, described in Section 2.2. Table 5 shows the experimental results. In addition, the numbers of training samples via three different context augmentation strategies are also reported. We observe that our exponential strategy achieves the significant performance compared with the linear and NOP strategies. With the increase of training samples, the performance grows consistently. Note that the number of training samples using our exponential strategy grows up to about 3.5 times on average. This would not cause too much computational burden.

Strategy	Rest 14		Lap 14		Rest 15		Rest 16	
	F1	#	F1	#	F1	#	F1	#
Exp	**72.01**	5258	**60.17**	3022	**64.53**	2044	**71.57**	2746
Linear	69.08	4102	56.52	2562	62.65	1724	69.37	2396
NOP	50.49	1266	50.40	906	51.25	605	55.95	857

Table 5: F1 scores and the number of training samples via three context augmentation strategies on $\mathcal{D}_2$.

Furthermore, as reported in Table 6, we observe that the increment in the multi-aspect setting is

[5]The AI and AAI stages contain $t + 1$ and t samples, respectively. The one produced by masking all aspects except the last duplicates. Hence, the number of total samples is $2t$.

Strategy	Rest 14		Lap 14		Rest 15		Rest 16	
	SA	MA	SA	MA	SA	MA	SA	MA
Exp	**73.48**	**71.31**	**60.44**	**59.86**	**64.83**	**64.09**	**69.86**	**73.66**
Linear	71.38	67.48	57.05	56.02	64.29	60.51	68.38	70.54
NOP	64.39	40.80	56.64	42.21	59.18	37.21	62.80	44.69

Table 6: F1 scores in the single-aspect (SA) and multi-aspect (MA) settings for different context augmentation strategies.

Full	**COM-MRC**	**67.07**
Module	*w/o* Aspect Representation	66.10 (0.97↓)
	w/o Opinion Representation	66.16 (0.91↓)
	w/o Existence Concatenation	66.62 (0.45↓)
	w/o Sentiment Attention	65.57 (1.50↓)

Table 7: The average F1 scores of ablation study on $\mathcal{D}_2$.

much larger than that in the single-aspect setting when compare Exp with other strategies. To sum up, the strategy of context augmentation in our COM-MRC is effective.

4.2 On Discriminative Model

Are the modules in our discriminative model effective? To this end, we conduct ablation experiments on $\mathcal{D}_2$. For aspect and opinion extraction modules, we remove the aspect representation and opinion representation in Figure 2, respectively. For the aspect detection module, we remove the concatenation in Eq. (12) using the *[CLS]* token representation h_{cls}. For the sentiment classification module, we remove the sentiment attention in Eq. (8) using only the context representation h_x. The experimental results are reported in Table 7. We observe that the sentiment attention has the largest impact, resulting in a 1.50% decrement on the performance. This shows our attention mechanism effectively fuse the semantic information within aspects and opinions. In addition, the performances of the other three variants decrease in some degree. To sum up, all four modules in our model contribute to the superior performance on ASTE task.

4.3 On Inference Method

Is our inference method effective? First, we show two versions of inference method. These two methods involve an identical AI stage but a different AAI stage. In Figure 3, we illustrate the two AAIs. The left denoted as AAI 1 is a naïve version which only masks necessary aspects. The right denoted as AAI 2 is the one used in our COM-MRC. In AAI 1, to obtain the opinions and the sentiment of each aspect, the aspects are masked one by one from left to right. However, the current aspect processing stage could be disturbed by the subsequent aspects. As shown in Figure 3, the aspect "*ambience*" processing stage is disturbed by "*place*". Thus, the term "*overrated*" is mistakenly identified as an opinion of "*ambience*".

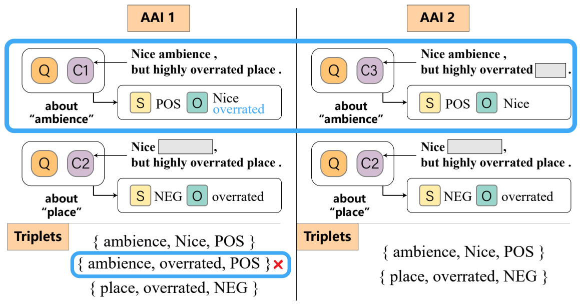


Figure 3: The comparison of the naïve AAI 1 (left) with AAI 2 adopted in our COM-MRC (right). The difference is highlighted with light blue boxes.

Mode	Inference	Rest 14	Lap 14	Rest 15	Rest 16
Single-Aspect	AAI 1/2	73.48	60.44	64.83	69.86
Multi-Aspect	AAI 1	68.80	57.13	60.32	71.50
	AAI 2	**71.31**	**59.86**	**64.09**	**73.66**
	Δ	+2.51	+2.73	+3.77	+2.16

Table 8: Comparison of F1 scores based on two AAIs on $\mathcal{D}_2$. The symbol Δ denotes the increment.

Furthermore, we conduct experiments using AAI 1 and AAI 2 based on COM-MRC framework. We consider both the single- and multi-aspect settings based on the discriminative model for $\mathcal{D}_2$. The experimental results are shown in Table 8. These two AAIs perform identically in the single-aspect setting. However, in the multi-aspect setting, AAI 2 outperforms AAI 1 significantly (a maximum increment 3.77% on Rest 15) on all four datasets. It indicates that other aspects would cause much interference and masking other aspects can alleviate the interference effectively. Note that the reason why not directly mask opinions is that an opinion may match multiple aspects and masking one opinion will blind the opinion extraction of all other aspects corresponding to it. Furthermore, the experimental results verify that our inference method in COM-MRC is effective especially for sentences with multiple aspects.

4.4 On Query

Is our query q effective? To answer this question, we conduct experiments with three types of queries. The first is the regular query adopted in our COM-MRC. The second is an improper query by removing the keyword "*first*". The third is null which means no query is provided. The experimental re-

Query	Rest 14	Lap 14	Rest 15	Rest 16
Regular Query: "*Find the first aspect term and corresponding opinion terms in the text*"	**72.01**	**60.17**	**64.53**	**71.57**
Improper Query: "*Find the aspect term and corresponding opinion terms in the text*"	70.44 (-1.57)	59.44 (-0.73)	62.97 (-1.56)	70.40 (-1.17)
Null	70.23 (-1.78)	57.86 (-2.31)	60.78 (-3.75)	69.02 (-2.55)

Table 9: F1 scores of different queries on $\mathcal{D}_2$.

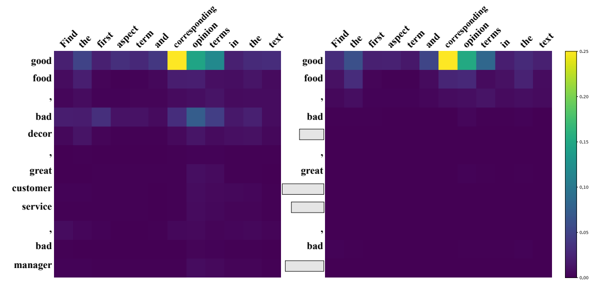
sults are reported in Table 9. The performance of the improper query decreases by a mean 1.26%. Compared with the improper query, a null query drops much more with a mean decrement of 2.60%. This shows the effectiveness of our query.

4.5 Attention Visualization

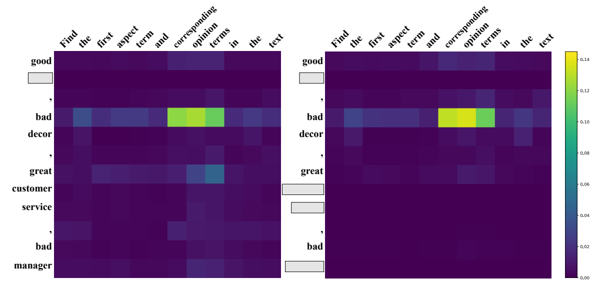
To show the effective treatment of the interference problem, we visualize the attention matrices, which imply the opinion information on the first aspect under the regular query and the masked context. Consider the sentence with two opposite polarities, "*good food, bad decor, great customer service, bad manager*". As shown in Figure 4(a) for identifying the opinion term of "*food*", both subfigures show that major attention is paid to the golden opinion "*good*". However, the left indicates there exists non-negligible attention on the incorrect opinions, especially on "*bad*". In contrast, the right shows if other aspect terms are masked, the attention on incorrect opinions could be drastically reduced. Similarly, Figure 4(b) shows the attention on incorrect opinions, especially on "*great*" can be reduced if other aspects are all masked. In addition, we observe that the span "*corresponding opinion terms*" in our query has high attention scores with golden opinions. To sum up, masking other aspects can effectively help identify current aspect information.

4.6 Case Study

In Table 10, we show several cases with multiple aspects to compare our COM-MRC with BMRC. In the first example, both methods correctly extract the aspect terms "*ambience*" and "*place*" with their opinion terms "*Nice*" and "*overrated*", respectively. However, BMRC fails to correctly identify the sentiment polarity of "*place*". Note that the sentiment polarity of "*ambience*" is also positive. In the second example, BMRC could extract a incorrect triplet. Here, two aspect terms, "*price*" and "*shipping*" do not share the opionion term "*great*".



(a) Visualization of attention matrices for the first aspect "*food*" under the query and two masked contexts.



(b) Visualization of attention matrices for the first aspect "*decor*". The attention head is the same as that in (a).

Figure 4: Visualization of attention matrices.

5 Related Work

ABSA generally comprises three subtasks: *Aspect Term Extraction* (ATE) (Hu and Liu, 2004; Yin et al., 2016; Li et al., 2018b; Xu et al., 2018; Ma et al., 2019; Chen and Qian, 2020; Wei et al., 2020), *Aspect Sentiment Classification* (ASC) (Wang et al., 2016b; Tang et al., 2016; Ma et al., 2017; Fan et al., 2018; Li et al., 2018a; Zhang et al., 2019; Sun et al., 2019; Wang et al., 2020; Li et al., 2021) and *Opinion Term Extraction* (OTE) (Yang and Cardie, 2012, 2013; Fan et al., 2019; Wu et al., 2020b). The studies ignore the correlations between these subtasks.

Some subsequent studies devoted to couple two subtasks. These works mainly are grouped into two tasks: *Aspect and Opinion Term Co-Extraction* (AOTE) (Wang et al., 2016a, 2017; Dai and Song, 2019; Wang and Pan, 2019; Chen et al., 2020; Wu et al., 2020a) and *Aspect-Sentiment Pair Extraction* (ASPE) (Ma et al., 2018; Li et al., 2019a,b; He et al., 2019). Most recently, ASTE as a new variant of ABSA has received wide attention. Peng et al. (2020) originally proposed a pipeline method that identifies aspects with sentiment and opinions independently then pairs them for forming sentiment triplets. End-to-end approaches (Xu et al., 2020; Wu et al., 2020a; Zhang et al., 2020; Chen et al., 2021b; Yan et al., 2021) are then proposed. Another mainstream framework is the MRC based

Sentence	Ground Truth	BMRC	Our COM-MRC
Nice ambience, but highly overrated place. (From Rest 16 of $\mathcal{D}_2$)	(ambience, Nice, POS) (place, overrated, NEG)	(ambience, Nice, POS) (place, overrated, POS) ✗	(ambience, Nice, POS) (place, overrated, NEG)
great price free shipping what else can i ask for!! (From Lap 14 of $\mathcal{D}_2$)	(price, great, POS) (shipping, free, POS)	(price, great, POS) (shipping, free, POS) (shipping, great, POS) ✗	(price, great, POS) (shipping, free, POS)

Table 10: Extraction results comparison of our COM-MRC with BMRC for the ASTE task.

methods (Mao et al., 2021; Chen et al., 2021a), whereas they are susceptible to interference from the existence of multiple aspect terms.

6 Conclusion and Future Work

In this paper, we propose a novel COntext-Masked MRC (COM-MRC) framework to alleviate the interference problem in the ASTE task. Our COM-MRC comprises three close-related components. The context augmentation can effectively expand the training corpus. The discriminative model comprising four modules works collaboratively. Our inference method involving two stages can effectively reduce the interference from other aspects. Extensive experiments on two groups of benchmark datasets demonstrate the effectiveness of our COM-MRC framework. In the future, we will devote to devising a one-stage method with a faster inference.

Limitations

In our COM-MRC framework, we design a context augmentation strategy. This strategy produces the masked context from 1 to 2^t for one sentence. This could increase the training samples achieving remarkable performance. On the other side, this would increase the training time of the discriminative model. Therefore, this would prevent our COM-MRC from applying to the scenarios with large-scale data.

Acknowledgements

This work was supported in part by the National Key R&D Program of China under Grant 2019YFF0303302 and in part by the National Natural Science Foundation of China under Grant 62076032. We appreciate constructive feedback from the anonymous reviewers for improving the final version of this paper.

References

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, Dublin, Ireland. 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.

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021a. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):12666–12674.

Zhexue Chen, Hong Huang, Bang Liu, Xuanhua Shi, and Hai Jin. 2021b. Semantic and syntactic enhanced aspect sentiment triplet extraction. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1474–1483, Online. Association for Computational Linguistics.

Zhuang Chen and Tieyun Qian. 2020. Enhancing aspect term extraction with soft prototypes. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2107–2117, 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.

Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3433–3442, Brussels, Belgium. Association for Computational Linguistics.

Zhifang Fan, Zhen Wu, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2019. Target-oriented opinion words extraction with target-fused neural sequence labeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2509–2518, Minneapolis, Minnesota. Association for Computational Linguistics.

Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier. 2019. An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 504–515, Florence, Italy. Association for Computational Linguistics.

Minghao Hu, Yuxing Peng, Zhen Huang, Dongsheng Li, and Yiwei Lv. 2019. Open-domain targeted sentiment analysis via span-based extraction and classification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 537–546, Florence, Italy. Association for Computational Linguistics.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 6319–6329, Online. Association for Computational Linguistics.

Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018a. 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. 2019a. A unified model for opinion target extraction and target sentiment prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):6714–6721.

Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. 2018b. Aspect term extraction with history attention and selective transformation. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 4194–4200. International Joint Conferences on Artificial Intelligence Organization.

Xin Li, Lidong Bing, Wenxuan Zhang, and Wai Lam. 2019b. Exploiting BERT for end-to-end aspect-based sentiment analysis. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, pages 34–41, Hong Kong, China. Association for Computational Linguistics.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Dehong Ma, Sujian Li, and Houfeng Wang. 2018. Joint learning for targeted sentiment analysis. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4737–4742, Brussels, Belgium. Association for Computational Linguistics.

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.

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, IJCAI'17, page 4068–4074. AAAI Press.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(15):13543–13551.

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. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):8600–8607.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris 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, Denver, Colorado. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 27–35, Dublin, Ireland. Association for Computational Linguistics.

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.

Duyu Tang, Bing Qin, and Ting Liu. 2016. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224, Austin, Texas. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

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 and Sinno Jialin Pan. 2019. Transferable interactive memory network for domain adaptation in fine-grained opinion extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):7192–7199.

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 Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, page 3316–3322. AAAI Press.

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.

Zhenkai Wei, Yu Hong, Bowei Zou, Meng Cheng, and Jianmin Yao. 2020. Don't eclipse your arts due to small discrepancies: Boundary repositioning with a pointer network for aspect extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3678–3684, Online. Association for Computational Linguistics.

Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020a. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2576–2585, Online. Association for Computational Linguistics.

Zhen Wu, Fei Zhao, Xin-Yu Dai, Shujian Huang, and Jiajun Chen. 2020b. Latent opinions transfer network for target-oriented opinion words extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):9298–9305.

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, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4755–4766, Online. 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.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. A unified generative framework for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2416–2429, Online. Association for Computational Linguistics.

Bishan Yang and Claire Cardie. 2012. Extracting opinion expressions with semi-Markov conditional random fields. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language*

Processing and Computational Natural Language Learning, pages 1335–1345, Jeju Island, Korea. Association for Computational Linguistics.

Bishan Yang and Claire Cardie. 2013. Joint inference for fine-grained opinion extraction. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1640–1649, Sofia, Bulgaria. Association for Computational Linguistics.

Yichun Yin, Furu Wei, Li Dong, Kaimeng Xu, Ming Zhang, and Ming Zhou. 2016. Unsupervised word and dependency path embeddings for aspect term extraction. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 2979–2985. AAAI Press.

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.

Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. 2020. A multi-task learning framework for opinion triplet extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 819–828, Online. Association for Computational Linguistics.