A Robustly Optimized BMRC for Aspect Sentiment Triplet Extraction

Shu Liu[1,2], **Kaiwen Li**[1,2], **Zuhe Li**[3,4,*]

[1]School of Computer Science and Engineering, Central South University
[2]Hunan Engineering Research Center of Machine Vision and Intelligent Medicine
[3]School of Computer and Communication Engineering, Zhengzhou University of Light Industry
[4]Henan Key Laboratory of Data Intelligence on Food Safety
{sliu35, itkaven}@csu.edu.cn   zuheli@126.com

Abstract

Aspect sentiment triplet extraction (ASTE) is a challenging subtask in aspect-based sentiment analysis. It aims to explore the triplets of aspects, opinions and sentiments with complex correspondence from the context. The bidirectional machine reading comprehension (BMRC) can effectively deal with ASTE task, but several problems remains, such as query conflict and probability unilateral decrease. Therefore, this paper presents a robustly optimized BMRC method by incorporating four improvements. The word segmentation is applied to facilitate the semantic learning. Exclusive classifiers are designed to avoid the interference between different queries. A span matching rule is proposed to select the aspects and opinions that better represent the expectations of the model. The probability generation strategy is also introduced to obtain the predicted probability for aspects, opinions and aspect-opinion pairs. We have conducted extensive experiments on multiple benchmark datasets, where our model achieves the state-of-the-art performance.[1]

1 Introduction

Aspect-based sentiment analysis (ABSA) is an important research area of natural language processing (NLP), which aims to mine fine-grained opinions and sentiments based on a specific aspect. In recent years, it has attracted extensive attention of researchers (Hu and Liu, 2004). ABSA includes three basic subtasks: aspect term extraction (Yin et al., 2016; Li et al., 2018; Ma et al., 2019), opinion term extraction (Liu et al., 2015; Wu et al., 2021), and aspect level sentiment classification (Wang et al., 2016; Chen et al., 2017; Jiang et al., 2019; Zhang and Qian, 2020).

Substantial progress has been achieved in recent studies, integrating multiple subtasks into a more complex task (Chen and Qian, 2020; He et al., 2019; Luo et al., 2019; Zhao et al., 2020). Among them, aspect sentiment triplet extraction (ASTE) (Peng et al., 2020) becomes a subject of great interest, which is also the goal of our work. Many research efforts have been made (Xu et al., 2021; Mao et al., 2021; Chen et al., 2021), for example, using bidirectional machine reading comprehension (BMRC) for ASTE. It handles the task effectively, but problems still remain. In the structure of BMRC, the shared classifiers may lead to query conflicts based on specific context, thus affecting the model performance. Some important strategies are also ignored, such as word segmentation, span matching and probability generation.

In this paper, we present a robustly optimized BMRC method for ASTE. The task is transformed into a machine reading comprehension problem. The complex correspondence between the aspect and opinion is processed through bidirectional query based on specific context. Such relationship can be effectively used to make their extraction mutually beneficial, thus facilitating better prediction of various sentiments. In order to deal with the ASTE task more efficiently, we incorporate the word segmentation and exclusive classifiers, and improve the span matching where the priority rule of the combination of probability and position relationship has been added. We also optimize the generation of probability to avoid its unilateral decrease. Our contributions can be summarized as follows:

- Exclusive classifiers are designed in BMRC, so as to avoid the interference between different question answering steps and the query conflict.

- We further advance the prediction performance by adding word segmentation, improving span matching and probability generation.

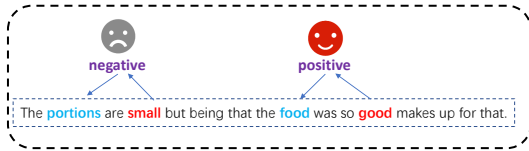- Extensive experiments are conducted on

---

Figure 1: The illustration of ASTE task.

benchmark datasets, where our model achieves the state-of-the-art performance.

2 Methodology

In this section, we briefly review the ASTE task and BMRC model, and then introduce our four improvements in detail.

2.1 Problem Formulation

Given a sentence $W = \{w_1, w_2, ..., w_M\}$ with M tokens, ASTE task is to identify the collection of triplets $T = \{(a_i, o_i, s_i)\}_{i=1}^{|T|}$, where a_i, o_i, s_i and $|T|$ denote the aspect, the opinion expression, the sentiment, and the number of triplets, respectively. For the sentence shown in Figure 1, the collection T is $\{(portions, small, negative), (food, good, positive)\}$.

2.2 BMRC

BMRC can put forward the corresponding query according to the context, and the model then outputs the desired answer.
Forward Query BMRC will query all the aspects based on context; Then, according to the aspect of each prediction, all opinions describing it are queried from the context.
Backward Query BMRC will query all the opinions based on context; Then, according to the opinion of each prediction, all aspects describing it are queried from the context.
Sentiment Prediction Once the aspect-opinion pairs are obtained, the sentiment queries can be constructed to predict the sentiments of the corresponding pairs according to the context.

After that, the sentiments and aspect-opinion pairs are combined into triplets. The whole process is illustrated in Figure 2.

2.3 Word Segmentation

We use the tokenizer based on wordpiece in BERT (Devlin et al., 2019) to segment words into subwords. Wordpiece is a common technique for word segmentation in NLP tasks.

The role of word segmentation has been investigated. Suppose the word *"walking"* is fed into the model, unless it appears many times in the training corpus, the model may fail to handle the word well. When similar words like *"walked"*, *"walker"* or *"walks"* show up, without word segmentation, they will be treated as completely different words. However, if they are subdivided into *"walk ##ing"*, *"walk ##ed"*, *"walk ##er"*, and *"walk ##s"*, their sub-word *"walk"* contains the same semantics which is quite common during training. In this sense, the model is able to learn more information through word segmentation.

2.4 Exclusive Classifiers

Bidirectional queries are performed in BMRC, and the model needs to perform multiple different types of queries based on context. For example, the aspect query in forward query is different from the opinion query in backward query. The former queries all the aspects in the context, while the latter queries all the opinions in the context, requiring different entities. Another example is the aspect query in the forward query and the aspect query in the backward query. Although the entities of the two queries are the same, the latter conveys opinion information and searches for all the aspects described by it, while the former does not carry any context information, namely, all the aspects in the context.

In the original BMRC, all queries share one classifier. However, if different types of queries use the same classifier, it cannot serve any part very well. These different types of queries will interfere with each other and cause the query conflict. By adding exclusive classifiers, each different type of query can use a unique classifier, as shown in Figure 3, which can effectively avoid the problem of query conflict and greatly improve the performance of the model.

2.5 Span Matching

Recently, there is a lot of work to deal with ABSA tasks based on span extraction (Hu et al., 2019; Xu et al., 2021), so does BMRC. After obtaining the predicted value of each position as the start or end position of span through binary classifiers, the predicted value is converted into probability using softmax function (Chen et al., 2021).

When predicting the span, many start and end positions may be predicted. The rule to match them is very important, which will seriously affect the
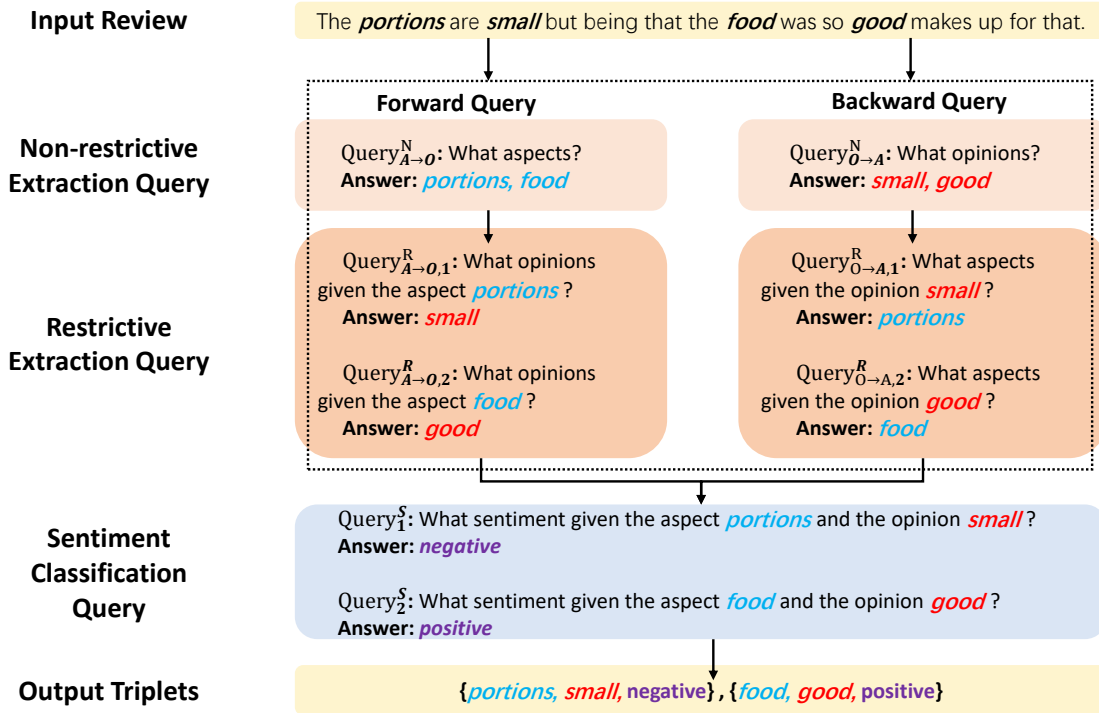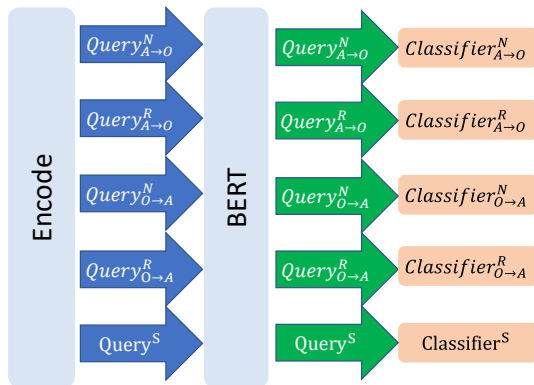
Figure 2: The BMRC framework.

Figure 3: Exclusive classifiers for different queries.

performance of the model. The matching should consider the probability of start and end positions, as well as the relationship between the positions. The former represents the optimistic degree of the model for the position, and the latter is the judgment that the start and end positions of span are as close as possible; the priority of probability is higher. So, the overview of our span matching rule is: make each end position match the start position with the highest probability after the previous end position. If there is a start position with the same probability, select the one whose position is closest to the end position.

2.6 Probability Generation

Once the bidirectional queries and span matching are completed, the aspects, opinions and pairs with corresponding relationship are obtained. In BMRC, the probability product of the start and end positions is taken as the probability of the span, and the probability of pair is the probability product of aspect and opinion. In this way, the probability of pair decreases unilaterally and cannot well represent the prediction of the pair by the model. For example, the probability of the four positions of pair is 0.9, while the probability of pair is $0.9^4 = 0.6561$, which seems not so reasonable.

By probability generation, we can effectively solve the problem of unilateral decrease in the probability of span and pair, so that their probability can better reflect the expectation of the model. The operations are shown in Equation 1 and 2, where we balance the probability of span and pair so that their probability is within the interval of the two related probabilities. It enables us to avoid the unilateral decrease of probability, but keep more appropriate to the expectation of the model.

$$P(span) = \sqrt{P(span_{start}) * P(span_{end})} \quad (1)$$

$$P(pair) = \sqrt{P(pair_{asp}) * P(pair_{opi})} \quad (2)$$

Span Matching

Forward Query

	This	place	is	a	must	visit	!
Start	5%	91%	2%	12%	5%	3%	0%
End	2%	95%	1%	1%	1%	3%	0%
Start	5%	10%	11%	21%	93%	3%	0%
End	2%	9%	1%	1%	2%	90%	0%

Backward Query

	This	place	is	a	must	visit	!
Start	1%	12%	10%	86%	86%	3%	1%
End	2%	15%	1%	1%	1%	87%	1%
Start	1%	90%	3%	6%	6%	3%	1%
End	2%	96%	1%	1%	1%	2%	1%

Probability Generation

$P(place)^{Aspect}_{A \to O} = 93\%$
$P(must\ visit)^{Opinion}_{A \to O} = 91\%$
$P(place, must\ visit)^{Pair}_{A \to O} = 92\%$

$P(must\ visit)^{Opinion}_{O \to A} = 86\%$
$P(place)^{Aspect}_{O \to A} = 93\%$
$P(place, must\ visit)^{Pair}_{O \to A} = 89\%$
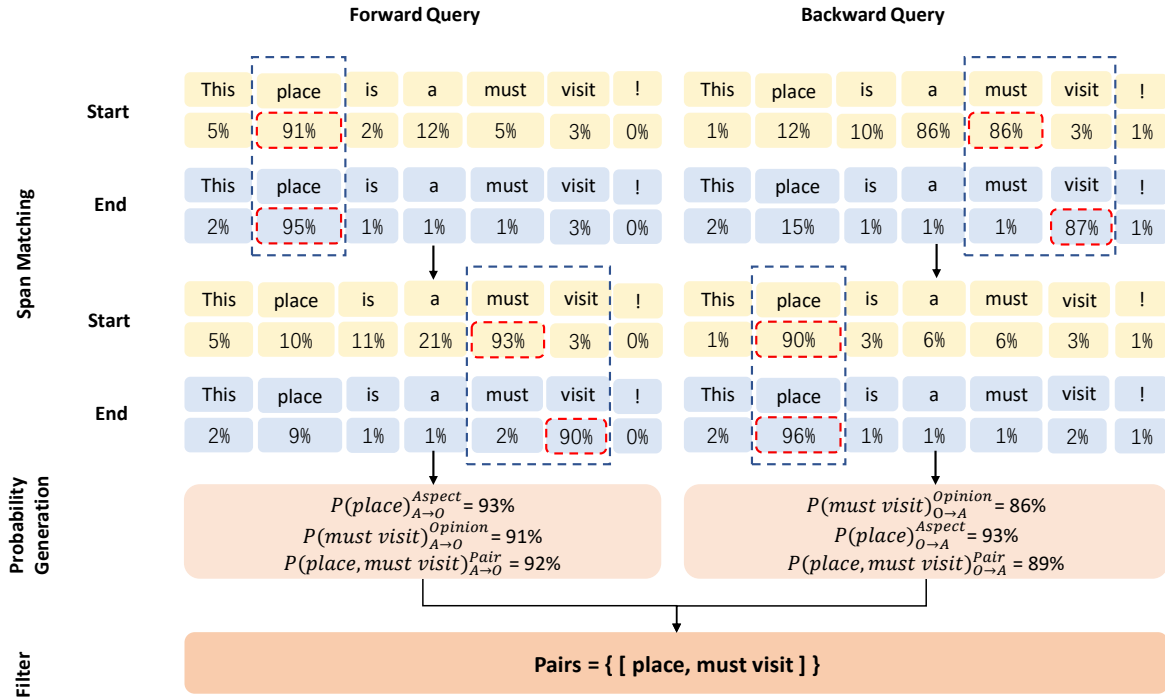
Filter

Pairs = { [place, must visit] }

Figure 4: Span matching and probability generation architecture.

In the above equations, $P(\star)$ represents the prediction probability of $\star$. The $span$ represents an aspect or opinion, $span_{start}$ and $span_{end}$ represent its start and end positions. The $pair$ represents an aspect-opinion pair, $pair_{asp}$ and $pair_{opi}$ represent the aspect and opinion in the pair. For the convenience of comparison, the calculation method of $P(span_{start})$ and $P(span_{end})$ is consistent with BMRC (Chen et al., 2021).

The effects of span matching and probability generation are shown in Figure 4.

3 Experiment

In this section, we introduce information about the experiments, including datasets, evaluation metrics, baselines, experimental results, and ablation study.

3.1 Datasets

We evaluate the model performance on ASTE-Data-v1 (Peng et al., 2020) and ASTE-Data-v2 (Xu et al., 2020), which are popular benchmark datasets for ASTE task. They are derived from Laptop14, Rest14, Rest15, and Rest16 of SemEval shared challenges (Pontiki et al., 2014, 2015, 2016). The ASTE-Data-v2 datasets are the refined data of the previous ASTE-Data-V1 datasets[2].

3.2 Evaluation Metrics

We use precision (P), recall (R) and F1 scores as evaluation metrics to gauge the performance. A triplet prediction is correct only if the aspect, opinion and sentiment are predicted correctly.

3.3 Results

We focus on the ASTE task. The experimental results on ASTE-Data-v1 and ASTE-Data-v2 datasets are shown in Tables 1 and 2, respectively. In order to make a fair comparison with baselines, our F1 scores appear at least three times in the experiments.

It is worth noting that we have achieved state-of-the-art performances on both ASTE-Data datasets, indicating that our improvements further advance the performance of BMRC in dealing with ASTE task. On the Laptop14, Rest14, Rest15, and Rest16 datasets of ASTE-Data-v1, the F1 scores of our model are increased by 2.97, 4.20, 5.61 and 5.52 respectively, compared with the original BMRC. As for ASTE-Data-v2, we also increase the F1 scores of the Strong baseline Span-ASTE (Xu et al., 2021) by 2.74, 0.77, 2.36 and 2.90, respectively. This indicates that our improvement is very significant.

Model	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Peng-two-stage (Peng et al., 2020)	44.18	62.99	51.89	40.40	47.24	43.50	40.97	54.68	46.79	46.76	62.97	53.62
JET-BERT (Xu et al., 2020)	67.97	60.32	63.92	58.47	43.67	50.00	58.35	51.43	54.67	64.77	61.29	62.98
Dual-MRC (Mao et al., 2021)	71.55	69.14	70.32	57.39	53.88	55.58	63.78	51.87	57.21	68.60	66.24	67.40
BMRC (Chen et al., 2021)	71.32	70.09	70.69	65.12	54.41	59.27	63.71	58.63	61.05	67.74	68.56	68.13
Span-BART (Yan et al., 2021)			72.46			57.59			60.11			69.98
Ours	**73.84**	**75.98**	**74.89**	**66.66**	**58.36**	**62.24**	**66.96**	**66.37**	**66.66**	**71.14**	**76.34**	**73.65**

Table 1: Experiments on the ASTE-Data-v1 datasets. The best P, R and F1 scores are in bold.

Model	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
JET-BERT (Xu et al., 2020)	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	**70.42**	58.37	63.83
Span-BART (Yan et al., 2021)	65.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
Span-ASTE (Xu et al., 2021)	**72.89**	70.89	71.85	63.44	55.84	59.38	62.18	64.45	63.27	69.45	71.17	70.26
Ours	72.51	**72.73**	**72.62**	**68.13**	**57.09**	**62.12**	**65.90**	**65.36**	**65.63**	69.98	**76.65**	**73.16**

Table 2: Experiments on the ASTE-Data-v2 datasets.

Improved Stack	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BMRC	73.42	66.70	69.89	69.53	52.11	59.57	64.06	55.87	59.69	70.37	66.53	68.39
+ Word Segmentation	72.89	68.71	70.74	72.70	51.01	59.95	64.30	59.79	61.96	**74.24**	67.31	70.61
+ Exclusive Classifiers	73.74	69.51	71.56	72.02	53.59	61.45	**66.51**	60.61	63.43	72.52	71.40	71.96
+ Span Matching	**74.65**	69.91	72.20	**73.11**	53.59	61.84	66.44	62.88	64.61	71.50	73.73	72.60
+ Probability Generation	72.51	**72.73**	**72.62**	68.13	**57.09**	**62.12**	65.90	**65.36**	**65.63**	69.98	**76.65**	**73.16**

Table 3: The performance of our four improvements on the ASTE-Data-v2 datasets.

3.4 Ablation Experiments

Firstly, we experiment the model without improvement on the ASTE-Data-v2. The model is a reproduction based on BMRC, and then gradually superimposes the four improvements of word segmentation, exclusive classifiers, span matching and probability generation to conduct ablation experiment. This arrangement corresponds to the sequence before and after they contact the data, that is, the data will first pass through word segmentation and enter the model, the prediction value is obtained from the exclusive classifiers, and then span matching is carried out according to it. Finally, the probability generation is used to generate probabilistic representations of aspects, opinions and pairs. The datasets and various parameters of the five experiments are the same. In order to make a fair comparison with baselines, our F1 scores appear at least three times in the experiments. The ablation experimental results are shown in Table 3. Each improvement advances the performance of the model, demonstrating their advantages and effectiveness.

4 Conclusion

In this paper, we propose several improvements on the basis of BMRC for ASTE task, which can effectively deal with the complex correspondence among aspect, opinion and sentiment. In order to deal with the problems of the original BMRC, we add exclusive classifiers and three strategies, including word segmentation, span matching and probability generation. The proposed method is expected to handle complex ASTE task more efficiently. Extensive experiments are conducted to demonstrate the advantages of our improvements.

References

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 452–461, Copenhagen, Denmark. Association for Computational Linguistics.

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

Zhuang Chen and Tieyun Qian. 2020. Relation-aware collaborative learning for unified aspect-based sentiment analysis. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3685–3694, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

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, Seattle, Washington, USA, August 22-25, 2004*, pages 168–177. ACM.

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.

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

Pengfei Liu, Shafiq Joty, and Helen Meng. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1433–1443, Lisbon, Portugal. Association for Computational Linguistics.

Huaishao Luo, Tianrui Li, Bing Liu, and Junbo Zhang. 2019. DOER: Dual cross-shared RNN for aspect term-polarity co-extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 591–601, Florence, Italy. Association for Computational Linguistics.

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. *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.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. 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.

Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. 2021. Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 3957–3963. International Joint Conferences on Artificial Intelligence Organization. Main Track.

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.

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 2016, New York, NY, USA, 9-15 July 2016*, pages 2979–2985. IJCAI/AAAI Press.

Mi Zhang and Tieyun Qian. 2020. Convolution over hierarchical syntactic and lexical graphs for aspect level sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3540–3549, Online. Association for Computational Linguistics.

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.