Abstract

Aspect Sentiment Triplet Extraction (ASTE) is a new fine-grained sentiment analysis task that aims to extract triplets of aspect terms, sentiments, and opinion terms from review sentences. Recently, span-level models achieve gratifying results on ASTE task by taking advantage of whole span predictions. However, all the spans generated by these methods inevitably share at least one token with some others, and these method suffer from the similarity of these spans due to their similar distributions. Moreover, since either the aspect term or opinion term can trigger a sentiment triplet, it is challenging to make use of the information more comprehensively and adequately. To address these concerns, we propose a span-level bidirectional cross-attention framework. Specifically, we design a similar span separation loss to detach the spans with shared tokens and a bidirectional cross-attention structure that consists of aspect and opinion decoders to decode the span-level representations in both aspect-to-opinion and opinion-to-aspect directions. With differentiated span representations and bidirectional decoding structure, our model can extract sentiment triplets more precisely and efficiently. Experimental results show that our framework significantly outperforms state-of-the-art methods, achieving better performance in predicting triplets with multi-token entities and extracting triplets in sentences with multi-triplets.

1 Introduction

Aspect-based sentiment analysis (ABSA) is an important field in natural language processing (NLP). The ABSA task contains various fundamental subtasks, such as aspect term extraction (ATE), opinion term extraction (OTE), and aspect-level sentiment classification (ASC). Recent studies focus on solving these tasks individually or doing a combination of two subtasks, such as aspect term polarity co-extraction (APCE), aspect opinion co-extraction (AOCE), and aspect-opinion pair extraction (AOPE). However, none of these sub-tasks aims to extract the aspect terms (AT) with their corresponding opinion terms (OT) and sentiment polarity (SP) simultaneously. To tackle this problem, [Peng et al., 2020] propose the aspect sentiment triplet extraction (ASTE) task which aims to extract (AT, OT, SP) triplets such as (hot dogs, top notch, positive) and (coffee, average, negative) in the example of Figure 1.

To solve the ASTE task, recent works [Peng et al., 2020; Wu et al., 2020; Mao et al., 2021] use sequential token-level methods and formulate this task as a sequence tagging problem. Although these works achieve competitive results, their token-level model suffer from cascading errors due to sequential decoding. Therefore, [Xu et al., 2021] propose a span-level model to capture the span-to-span interactions among ATs and OTs by enumerating all possible spans as input. However, enumerating all possible spans inevitably causes each span to have other spans with which it shares tokens. For example, the aspect span hot dogs shares tokens with The hot dogs, The hot, hot dogs are, and so on. These similar spans may have adjacent distributions in the feature spaces, which lead to the false prediction of downstream tasks.

Besides, although bidirectional predicting sentiment triplets has been proved to be effective in ASTE task, existing bidirectional approaches [Chen et al., 2021] have a few drawbacks. Although these works propose a bidirectional frame-
work to identify aspect sentiment triplets in both aspect-to-opinion and opinion-to-aspect directions, they fail to classify sentiments for triplets in parallel, which ignores the interdependence and indicative association among different triplet candidates. For example, the triplets \{coffee, average, negative\} can help infer the sentiment polarity of \{coffee, top notch\} pair during sentiment classification because average and top notch express opposite sentiment tendencies. Apart from that, existing bidirectional approaches are mostly based on token-level models, which suffer from cascading errors due to sequential decoding.

In this paper, we propose a span-level bidirectional cross-attention (SBC) framework for ASTE task. Similar to prior span-level works [Lee et al., 2017; Zhao et al., 2020; Zhong and Chen, 2020; Dixit and Al-Onaizan, 2019], our framework enumerates all possible spans as the input. To separate the share-token spans which have adjacent distributions in feature space, we design a similar span separation loss to maximize the KL divergence of the distributions among similar spans. Based on the differentiated span representations, we further design a bidirectional cross-attention structure consist of an aspect decoder and an opinion decoder to identify triplets in both aspect-to-opinion and opinion-to-aspect directions, as shown in Figure 1. In the aspect-to-opinion direction, the aspect decoder aims to extract ATs such as \{hot dogs, coffee\}, and the opinion decoder aims to extract OTs such as \{top notch\} for each specific AT like \{hot dogs\}. Analogously, in the opinion-to-aspect direction, the opinion decoder and aspect decoder are aim to extract OTs and their corresponding ATs, respectively. In each direction, the sentiments are classified both in AT and OT extractions, and the sentiment of triplet is determined based on the confidence scores. To verify the effectiveness of our framework, we conduct a series of experiments based on four benchmark datasets. The experimental results show our framework substantially outperforms the existing methods. In summary, our contributions are as follows:

- We design a span-level bidirectional cross-attention framework to identify aspect sentiment triplets in both aspect-to-opinion and opinion-to-aspect directions.
- We propose the similar span separation loss to separate the representations of spans which share same tokens. Based on these differentiated span representations, the bidirectional cross-attention structure can extract the sentiment triplets more precisely.
- Our proposed framework not only achieves state-of-the-art performance in ASTE task, but better results in multi-token span and multi-triplet sentence scenarios.

3 Methodology

As shown in Figure 2, our SBC framework consists of five parts: task definition, span generation, similar span separation loss, bidirectional cross-attention structure, and inference. The details of all parts are given in the following subsections.

3.1 Task Definition

Given a sentence $S = \{w_{1}, w_{2}, \ldots, w_{n}\}$ consisting $n$ words, the goal of the ASTE task is to extract a set of aspect sentiment triplets $T = \{(a, o, c)_{k}\}_{k=1}^{T}$ from the given sentence $S$, where $(a, o, c)$ refers to (aspect term, opinion term, sentiment polarity) and $c \in \{Positive, Neutral, Negative\}$.

3.2 Span Generation

Given a sentence $S$ with $n$ tokens, there are $m$ possible spans in total. Each span $s_{i} = \{w_{\text{start}(i)}, \ldots, w_{\text{end}(i)}\}$ is defined by all the tokens from $\text{start}(i)$ to $\text{end}(i)$ inclusive, and the maximum length of span $s_{i}$ is $l$:

$$1 \leq \text{start}(i) \leq \text{end}(i) \leq n \quad (1)$$

To obtain span representations, we need to get the token-level representations first. In this paper, we utilize BERT [Devlin et al., 2018] as a sentence encoder to obtain token-level contextualized representations $\{h_{1}, h_{2}, \ldots, h_{n}\}$ of the given sentence $S$. Then, the token-level representations are combined by max pooling. Note that various methods can be applied to generate the representations for spans, the effectiveness of these span generation methods will be investigated in the ablation study in Appendix. We define the representation of span $s_{i}$ as:

$$g_{i} = \text{Max} (h_{\text{start}(i)}, h_{\text{start}+1(i)}, \ldots, h_{\text{end}(i)}) \quad (3)$$

where $\text{Max}$ represents max pooling.
3.3 Similar Span Separation Loss

After generating the representation of span, most previous models directly use the span representations for downstream tasks. However, enumerating all possible spans in a sentence inevitably generates lots of spans that have same tokens with some others, and the model may suffer from the limitations in processing these similar spans due to their adjacent distribution. To separate these spans with similar distributions, we propose a similar span separation loss function based on KL divergence for separating spans with shared tokens, as shown in Figure 2. The similar span separation loss is defined as:

\[ KL(G_i||G_j) = \sum_i \frac{\text{softmax}(g_i) \log \text{softmax}(g_j)}{\text{softmax}(g_j)} \]  

\[ KL(G_j||G_i) = \sum_j \frac{\text{softmax}(g_j) \log \text{softmax}(g_i)}{\text{softmax}(g_i)} \]  

\[ J_{KL} = \sum_i \log(1 + \frac{2}{KL(G_i||G_i) + KL(G_i||G_j)}) \]  

where \( G_i \) indicates the set of the representation of spans which share at least one token with \( s_i \).

3.4 Bidirectional Cross-attention Structure

As the aspect sentiment triplet can be triggered by an AT or an OT, we further design a bidirectional cross-attention structure to decode the span representations. As shown in Figure 2, the bidirectional cross-attention structure consists of an aspect decoder and an opinion decoder. The details of each component of bidirectional cross-attention structure are given in the following subsections.

**Aspect-to-opinion Direction**

In aspect-to-opinion direction (Blue arrows and modules in Figure 2), at first, the aspect decoder aims to extract all ATs along with their sentiment from the sentence. We can obtain the confidence score as well as the probability of the sentiment of AT as follows:

\[ u_i^a = FFN_A(g_i, \theta_a) \]  

\[ q_i^{a\rightarrow o,a} = w_{a\rightarrow o,a} u_i^a \]  

\[ p_i^{a\rightarrow o,a} = \text{softmax}(q_i^{a\rightarrow o,a}) \]  

where \( FFN_NA \) represents the FFNN of aspect decoder, \( \theta_a \) is the parameter for the FFNN, \( w_{a\rightarrow o,a} \in \mathbb{R}^{m \times c} \) is a trainable weight vector, and \( c^* \in \{ \text{Positive, Neutral, Negative, Null} \} \) is the number of sentiment polarity, \( \text{Null} \) here means that the corresponding span is not a valid AT.

After that, giving a set \( G_a \) of original span representations of all valid ATs \( g_j^a \in G_a \), we apply the opinion decoder to identify all OTs along with their sentiment for each particular valid AT by exploiting attention mechanism. Similarly, we obtain the probability distribution of the OT’s sentiment along with its confidence score by:

\[ u_i^o = FFN_O(g_i, \theta_o) \]  

\[ o_{i,j}^{a\rightarrow o} = \exp(u_i^o) \]  

\[ q_{i,j}^{a\rightarrow o,o} = w_{a\rightarrow o,o} (u_i^o + o_{i,j}^{a\rightarrow o} \cdot g_j^o) \]  

\[ p_{i,j}^{a\rightarrow o,o} = \text{softmax}(q_{i,j}^{a\rightarrow o,o}) \]  

where \( FFN_O \) represents the FFNN of opinion decoder, \( \theta_o \) is the parameter for the FFNN, \( w_{a\rightarrow o,o} \in \mathbb{R}^{m \times c} \) is a trainable weight vector. Furthermore, define the loss of aspect-to-opinion direction as:

\[ J_{a\rightarrow o} = -\sum_i y_i^{a\rightarrow o,a} \log(q_i^{a\rightarrow o,a}) \]  

\[ -\sum_i \sum_j y_{i,j}^{a\rightarrow o,o} \log(q_{i,j}^{a\rightarrow o,o}) \]  

where \( y_i^{a\rightarrow o,a} \) and \( y_{i,j}^{a\rightarrow o,o} \) are ground truth labels of the sentiments for AT and OT given a specific valid AT, respectively.
Opinion-to-aspect Direction
As for opinion-to-aspect direction (Red arrows and modules in Figure 2), the opinion decoder is deployed first to extracts all the OTs along with their sentiment from the sentence. To minimize the number of model parameters, the opinion decoder in both aspect-to-opinion and opinion-to-aspect directions shares the FFNN features, as described in Equation (10). The probability distribution of the sentiments of OTs as well as the confidence scores can be obtained as:

\[ q_{ij}^{a,o} = w_{a,o} u^a_i \] (15)

\[ p_{ij}^{a,o} = \text{softmax}(q_{ij}^{a,o}) \] (16)

where \( w_{a,o} \in \mathbb{R}^{m \times c} \) is a trainable weight vector. Given a set \( G_o \) of original span representations of all valid OTs \( x^{o}_{ij} \in G_o \), the aspect decoder is deployed to identify the ATs and their sentiment for each particular valid OT. Note that the aspect decoder in opinion-to-aspect direction also shares same FFNN features described in Equation (7) with the aspect decoder in aspect-to-opinion direction. The logits of ATs and their confidence scores in opinion-to-aspect direction can be obtained by:

\[ \alpha_{ij}^{a} = \frac{\exp(u^a_i)}{\exp(g^o_j)} \] (17)

\[ q_{ij}^{a,o} = w_{a,o,a} (u^a_i + \alpha_{ij}^{a} \cdot g^o_j) \] (18)

\[ p_{ij}^{a,o} = \text{softmax}(q_{ij}^{a,o}) \] (19)

where \( w_{a,o,a} \in \mathbb{R}^{m \times c} \) is a trainable weight vector. Finally, the loss for opinion-to-aspect direction is defined as:

\[ \mathcal{J}_{o \rightarrow a} = - \sum_{i} y_{i}^{a,o} \log (q_{i}^{a,o}) \] (20)

\[ - \sum_{i} y_{i}^{a,o} \log (p_{i}^{a,o}) \]

where \( y_{i}^{a,o} \) and \( p_{i}^{a,o} \) are the ground truth labels. Then, we combine the above loss functions to form the loss objective of the entire model:

\[ \mathcal{J} = \mathcal{J}_{KL} + \mathcal{J}_{a \rightarrow o} + \mathcal{J}_{o \rightarrow a} \] (21)

3.5 Inference
During training, the ground truth of all ATs, OTs, and their corresponding sentiment polarities are already known. Therefore, our model does not form the triplets during the training process. However, in the inference process of each direction, our model identify the triples in a pipeline. For more precise description of the inference process, we propose a inference algorithm to show the determination of the sentiment of each triplet and the combination of the triplet results from two directions. As illustrated in Algorithm 1, the final sentiment polarity of each triplet is determined based on the confidence scores of the corresponding sentiments in AT and OT extraction. And the final aspect sentiment triplets are the concatenation of the triplets in both aspect-to-opinion and opinion-to-aspect directions.

Algorithm 1 Inference Algorithm for ASTE of the SBC Framework

Input: sentence \( S \)

Output: triplets \( T = \{ (a, o, c) \} \)

Initialize \( T = \{ \} \);

Use \( S \) to generate all possible span representations \( G = \{ g_1, g_2, \ldots, g_m \} \) described in Section 3.2;

for \((\beta, \gamma) \in [(\text{aspect}, \text{opinion}), (\text{opinion}, \text{aspect})]\) do

Input \( G \) to the \( \beta \) decoder described in Section 3.4, output the valid \( \beta \) candidates \( \beta_i \in G_\beta \), the corresponding sentiment \( c^{\beta}_{ij} \) and the score of each sentiment \( q^{\beta}_{ij} \);

for \( \beta_i \in G_\beta \) do

Input \( G \) and the span representation \( g^{\beta}_{ij} \) of valid \( \beta \) candidates \( \beta_i \in G_\beta \) to the \( \gamma \) decoder, output the valid \( \gamma \) candidates \( \gamma_{ij} \) of the given span \( \beta_i \), the corresponding sentiment \( c^{\gamma}_{ij} \) and the score of each sentiment \( q^{\gamma}_{ij} \);

if \( q^{\beta}_{ij} > q^{\gamma}_{ij} \) then

triplet sentiment \( c = c^{\beta}_{ij} \)

else

triplet sentiment \( c = c^{\gamma}_{ij} \)

end if

\( T \leftarrow T \cup (\beta_i, \gamma_{ij}, c) \)

end for

end for

4 Experiments

4.1 Datasets

To verify the effectiveness of our proposed SBC framework, we conduct experiments on four benchmark datasets [Xu et al., 2020], which are constructed based on the original SemEval ABSA Challenges and the datasets of [Fan et al., 2019]. Table 1 lists the statistics of these datasets.

4.2 Experimental Setting

We adopt the cased base version of BERT [Devlin et al., 2018] in our experiments, which contains 110M parameters. During training, we use AdamW [Loshchilov and Hutter, 2017] to optimize the model parameters. The fine-tuning rate for BERT and the learning rate for other models are set to 1e-5 and 1e-4 respectively. Meanwhile, the mini-batch size is set to 6 and the dropout rate is set to 0.1. The maximum length of generated spans is set to 8. We train our framework in a total of 120 epochs on a NVIDIA Tesla V100 GPU.

4.3 Evaluation

To comprehensively evaluate the performance of different methods, we use precision, recall, F1-score as the evaluation metrics. The extracted ATs and OTs are considered correct if and only if predicted spans exactly match the ground truth spans. In the experiment, we select the testing results when the model achieves the best performance on the development set.

4.4 Baselines

To demonstrate the effectiveness of SBC framework, we compare our method with the following baselines:
or the OTs are multi-word spans. ‘#MW’ denotes the numbers of triplets that at least one of the ATs or OTs are single word spans. ‘#SW’ denotes the numbers of triplets where the ATs and OTs are single word spans.

Table 1: Statistics of the datasets. ‘#S’ denote the numbers of sentence, ‘POS’, ‘NEU’, and ‘NEG’ denote the numbers of positive, neutral, and negative triplets respectively. ‘#SW’ denotes the numbers of triplets where the ATs and OTs are single word spans. ‘#MW’ denotes the numbers of triplets that at least one of the ATs or the OTs are multi-word spans.

| Datasets | #S | POS | NEU | NEG | #SW | #MW |
|----------|----|-----|-----|-----|-----|-----|
| 14LAP    | 1206 | 1692 | 166 | 480 | 1586 | 752 |
| Train    | 310  | 404  | 54  | 119 | 388  | 189 |
| Test     | 492  | 773  | 63  | 155 | 657  | 337 |
| 14RES    | 906  | 817  | 126 | 517 | 824  | 636 |
| Train    | 219  | 169  | 36  | 141 | 190  | 156 |
| Test     | 328  | 364  | 63  | 116 | 291  | 252 |
| 15RES    | 605  | 783  | 25  | 205 | 678  | 335 |
| Train    | 148  | 185  | 11  | 53  | 165  | 84  |
| Test     | 322  | 317  | 25  | 143 | 297  | 188 |
| 16RES    | 857  | 1015 | 50  | 329 | 918  | 476 |
| Train    | 210  | 252  | 11  | 76  | 216  | 123 |
| Test     | 326  | 407  | 29  | 78  | 344  | 170 |

- **Peng-two-stage** [Peng et al., 2020] is a two-stage pipeline model. Peng-two-stage extracts both aspect-sentiment pairs and opinion terms in the first stage. In the second stage, Peng-two-stage pairs up the extraction results into triplets via an relation classifier.

- **JET** [Xu et al., 2020] is an end-to-end model which proposes a novel position-aware tagging scheme to jointly extracting the triplets. It also designs factorized feature representations so as to effectively capture the interaction among the triplet factors.

- **GTS** [Wu et al., 2020] is an end-to-end model which formulates ASTE as a unified grid tagging task. It first extracts the sentiment feature of each token, and then gets the initial prediction probabilities of token pairs based on these token-level features. It also designs an inference strategy to exploit the potential mutual indications between different opinion factors and performs the final prediction.

- **Dual-MRC** [Mao et al., 2021] is a joint training model which consists of two machine reading comprehensions. One of the MRC is for aspect term extraction, and another is for aspect-oriented opinion term extraction and sentiment classification.

- **B-MRC** [Chen et al., 2021] formalizes the ASTE task as a multi-turn machine reading comprehension task, and proposes three types of queries to extract targets, opinions and the sentiment polarities of aspect-opinion pairs, respectively.

- **Span-ASTE** [Xu et al., 2021] considers all possible spans in a sentence to consider the interaction between the whole spans of aspect terms and opinion terms when predicting their sentiment relation. They also propose a dual-channel span pruning strategy to ease the high computational cost caused by span enumeration.

### 4.5 Main Results

Table 2 reports the results of our framework and baseline models. According to the results, our framework achieves state-of-the-art performance on all datasets. Specifically, our framework surpasses the best baselines by an average of 2.3 F1-score on ASTE. This result demonstrates that our framework can distinguish the representation distributions between similar spans and take advantage of bidirectional decoding. Although some of the precision scores are slightly lower than B-MRC, the increase in recall significantly outperforms the previous baselines in most datasets, which shows the higher prediction completeness of our framework. It is worth noting that BMRC and Dual-MRC achieve better performance than JET and Peng-two-stage. This is probably because BMRC and Dual-MRC formalize the ASTE task as a multi-turn machine reading comprehension task and benefit from asking the model questions. Besides, we also observe that Span-based methods (tagging the start/end positions) show superior performance to sequence tagging methods. This is probably because sequence tagging methods extract entities by determining the label for every token, by which the compositionality of candidate labels is higher than span-based methods. Unlike those approaches, Span-ASTE and our method both utilize the span-level interactions to handle the ASTE task and avoid the cascading errors. Moreover, although Span-ASTE and our method both use span-level model to extract the triplets, our model still outperforms Span-ASTE. This is mainly because instead of matching each aspect span and opinion span, our method identify the triplets from both aspect-to-opinion and opinion-to-aspect directions and utilize Kullback-Leibler divergence to disperse the distribution among similar spans.

### 4.6 Ablation Study

To examine the effect of KL Loss, bidirectional cross-attention structure and the inference algorithm, we conduct an ablation study on 14LAP and 14RES datasets. As shown in Table 3, in most scenarios, the performance of $C_o$ are better than $C_a$, it is probably because classifying the sentiments of OTs is easier than classifying that of ATs. Although $C_o$ shows inferior performance in SBC$_{A→O}$ model, most of its results are on par with or even surpass other options, which indicates the effectiveness of our sentiment determination algorithm. Moreover, SBC$_{A→O}$ shows superior performance than SBC$_{O→A}$ on 14LAP datasets, while SBC$_{O→A}$ shows better performance on 14RES. SBC, however, outperforms other two unidirectional models on both 14LAP and 14RES datasets. This clearly indicates the effectiveness of the bidirectional cross-attention structure on decoding the span representations. The comparison results between applying kl loss and without kl loss show that magnifying the kl divergence among spans that share same tokens is beneficial to span representation and can contribute to the performance of downstream triplets extraction in both directions.

### 4.7 Effect of Entity Length

To investigate the performance of different methods on the ATE and OTE with different entity lengths, we report the F1 scores of our framework, Span-ASTE, GTS, and B-MRC on the extraction task with different lengths of entities. The results are illustrated in Figure 3. As the entity length increases,

1Refer to Appendix for the ablation study on the method of generating span representations and the complexity analysis on our framework and baselines.
Table 2: Precision (%), Recall (%) and F1 score (%) on the test set of the ASTE tasks. State-of-the-art results are marked bold. * indicates that the result is reproduced by us.

| Method         | 14LAP  | 14RES  | 15RES  | 16RES  |
|----------------|--------|--------|--------|--------|
|                | P      | R      | F1     | P      | R      | F1     | P      | R      | F1     |
| PENG-two-stage | 40.40  | 47.24  | 43.50  | 44.18  | 54.68  | 46.79  | 46.76  | 62.97  | 53.62  |
| JETi           | 51.48  | 42.65  | 46.65  | 50.00  | 67.97  | 60.32  | 64.77  | 61.29  | 62.98  |
| JETO           | 58.47  | 43.67  | 50.00  | 58.35  | 51.43  | 54.67  | 64.77  | 61.29  | 62.98  |
| GTS-BERT       | 57.52  | 51.92  | 54.58  | 70.92  | 69.49  | 70.20  | 59.29  | 58.07  | 58.67  |
| Dual-MRC       | 57.39  | 55.88  | 55.58  | 71.55  | 69.14  | 70.32  | 63.78  | 51.87  | 57.21  |
| B-MRC          | 70.89* | 50.20* | 58.78* | 75.41* | 64.04* | 69.26* | 69.83* | 56.04* | 58.74* |
| Span-ASTE      | 63.44  | 55.84  | 59.38  | 72.89  | 70.89  | 71.85  | 62.18  | 64.45  | 63.27  |
| SBC            | 63.64  | 61.80  | 62.71  | 77.09  | 70.99  | 73.92  | 63.00  | 64.95  | 63.96  |

Table 3: Experimental results of the ablation study on the KL Loss, inference algorithm and the bidirectional cross-attention structure (F1-score, %). $C_s$, $C_a$, and $C_o$ denote determining the sentiments of triplets based on confidence scores, on the sentiment of ATs, and the sentiment of OTs, respectively.

| Method | 14LAP | 14RES |
|--------|-------|-------|
|        | $C_s$ | $C_a$ | $C_o$ |
| SBC    | 62.71 | 62.14 | 62.38 |
| -kl loss | 61.77 | 61.37 | 61.48 |
| SBC$_{A\rightarrow O}$ | 62.41 | 61.35 | 62.60 |
| -kl loss | 60.88 | 60.08 | 60.92 |
| SBC$_{O\rightarrow A}$ | 62.26 | 61.84 | 61.84 |
| -kl loss | 60.70 | 60.70 | 60.70 |

Table 4: Effects of multiple triplets in a sentence in 14LAP for ASTE task (F1-score, %).

4.8 Effect of Multiple Triplets

To further verify the ability of our framework to handle multiple triplets, we compare the performance of our SBC framework and other baselines on ASTE task with different triplet numbers, and the results are shown in Table 4. We divide the sentences in 14LAP test set into 5 subclasses. Each subclass contains sentences with 1, 2, 3, 4, or $\geq 5$ triplets, respectively. When extracting triplets from sentences that contain 1 or 2 triplets, the performance of our framework is competitive to other models. However, when the number of triplets increases, the performance of Span-ASTE, GTS, and B-MRC decrease significantly, while the performance of our SBC framework remains stable or even slightly increases. These experimental results demonstrate the efficiency and stability of our framework in handling multiple triplets in a sentence.

5 Conclusions

In this work, we propose a span-level bidirectional cross-attention (SBC) framework for ASTE tasks. This span-level model can take advantages from both aspect-to-opinion and opinion-to-aspect directions and separate the spans with same tokens. The bidirectional decoding can ensure that either an AT or an OT can trigger an aspect sentiment triplet, which is more in line with human perception. For the shortage that spans which share at least one token may have adjacent representations, the similar span separation loss is deployed to maximize the KL divergence of these representations. The experimental results demonstrate that our SBC framework significantly outperforms the compared baselines and achieves state-of-the-art performance. Our bidirectional cross-attention structure has been proved to be effective in Section 4.6, but how to elegantly combine the extraction results in both directions can be a challenge for further investigation.
References

[Chen et al., 2021] Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. arXiv preprint arXiv:2103.07665, 2021.

[Devlin et al., 2018] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

[Dixit and Al-Onaizan, 2019] Kalpit Dixit and Yaser Al-Onaizan. Span-level model for relation extraction. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5308–5314, 2019.

[Fan et al., 2019] Zhifang Fan, Zhen Wu, Xinyu Dai, Shujian Huang, and Jiajun Chen. 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, 2019.

[Gao et al., 2021] Lei Gao, Yulong Wang, Tongcun Liu, Jingyu Wang, Lei Zhang, and Jianxin Liao. Question-driven span labeling model for aspect–opinion pair extraction. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 35, pages 12875–12883, 2021.

[Hu et al., 2018] Mengting Hu, Shiwang Zhao, Li Zhang, Keke Cai, Zhong Su, Renhong Cheng, and Xiaowei Shen. Can: Constrained attention networks for multi-aspect sentiment analysis. arXiv preprint arXiv:1812.10735, 2018.

[Lee et al., 2017] Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer. End-to-end neural coreference resolution. In Martha Palmer, Rebecca Hwa, and Sebastian Riedel, editors, Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP 2017, Copenhagen, Denmark, September 9-11, 2017, pages 188–197. Association for Computational Linguistics, 2017.

[Li et al., 2019a] Xin Li, Lidong Bing, Piji Li, and Wai Lam. A unified model for opinion target extraction and target sentiment prediction. In The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019. The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019. The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, IAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019, pages 6714–6721. AAAI Press, 2019.

[Li et al., 2019b] Zheng Li, Ying Wei, Yu Zhang, Xiang Zhang, and Xin Li. Exploiting coarse-to-fine task transfer for aspect-level sentiment classification. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 33, pages 4253–4260, 2019.

[Loshchilov and Hutter, 2017] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam. CoRR, abs/1711.05101, 2017.

[Ma et al., 2019] Dehong Ma, Sujian Li, Fangzhao Wu, Xing Xie, and Houfeng Wang. 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, 2019.

[Mao et al., 2021] Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. A joint training dual-mrc framework for aspect based sentiment analysis. In Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, pages 13543–13551. AAAI Press, 2021.

[Peng et al., 2020] Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 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, 2020.

[Wu et al., 2020] Zhen Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. Grid tagging scheme for aspect-oriented fine-grained opinion extraction. CoRR, abs/2010.04640, 2020.

[Xu et al., 2020] Lu Xu, Hao Li, Wei Lu, and Lidong Bing. Position-aware tagging for aspect sentiment triplet extraction. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020, pages 2339–2349. Association for Computational Linguistics, 2020.

[Xu et al., 2021] Lu Xu, Yew Ken Chia, and Lidong Bing. Learning span-level interactions for aspect sentiment triplet extraction. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021, pages 4755–4766. Association for Computational Linguistics, 2021.

[Yu et al., 2019] Jianfei Yu, Jing Jiang, and Rui Xia. Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. IEEE ACM Trans. Audio Speech Lang. Process., 27(1):168–177, 2019.

[Zhao et al., 2020] He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and Hui Xue. Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In Dan Jurafsky, Joyce Chai, Natalie Schuler, and Joel R. Tetreault, editors, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020, pages 3239–3248. Association for Computational Linguistics, 2020.

[Zhong and Chen, 2020] Zexuan Zhong and Danqi Chen. A frustratingly easy approach for joint entity and relation extraction. CoRR, abs/2010.12812, 2020.