



Enhancing Conversational Aspect-Based Sentiment Quadruple Analysis with Context Fusion Encoding Method

Xisheng Xiao, Jiawei Chen, Qianer Li, Peijie Huang^(✉), and Yuhong Xu

College of Mathematics and Informatics, South China Agricultural University,
Guangzhou, China
pjhuang@scau.edu.cn

Abstract. Aspect-based sentiment analysis (ABSA) has been a hot research topic due to its ability to fully exploit people's opinions through social media texts. Compared with analyzing sentiment in short texts, conversational aspect-based sentiment analysis, also known as DiaASQ, aiming to extract the sentiment quadruple of target-aspect-opinion-sentiment in a dialogue, is a relatively new task that involves multiple speakers with varying stances in a conversation. Conversations are longer than ordinary texts and have richer contexts, which can lead to context loss and pairing errors. To address this issue, this work proposes a context-fusion encoding method based on conversation threads and lengths to integrate the speech of different speakers, enabling the model to better understand conversational context and extract cross-utterance quadruples. Experimental results have demonstrated that the proposed method achieves an average F1-score of 42.12% in DiaASQ, which is 6.48% higher than the best comparative model, indicating superior performance.

Keywords: Conversational aspect-based sentiment quadruple analysis · Context fusion · Conversational opinion mining

1 Introduction

In recent years, many people have shared their opinions and reviews on the internet through various social media platforms. Fully mining information from these texts can provide significant help in improving products and increasing efficiency, making aspect-based sentiment analysis (ABSA) [1–3] a popular research direction. ABSA is a task that aims to detect fine-grained sentiment towards specific aspects of targets. Initially, ABSA only focused on aspect terms and sentiment polarities [4–6]. Later, researchers gradually realized the importance of two other key factors that influence sentiment polarity judgments: opinion terms and categories [2]. Depending on the different elements of interest, various ABSA tasks have been proposed, including pair extraction tasks (e.g. aspect-opinion pair extraction, AOPE [7]), triple ABSA tasks (e.g. aspect sentiment triplet extraction, ASTE [8, 9]), and quadruple ABSA tasks (e.g. aspect sentiment quad prediction, ASQP [10, 11]).

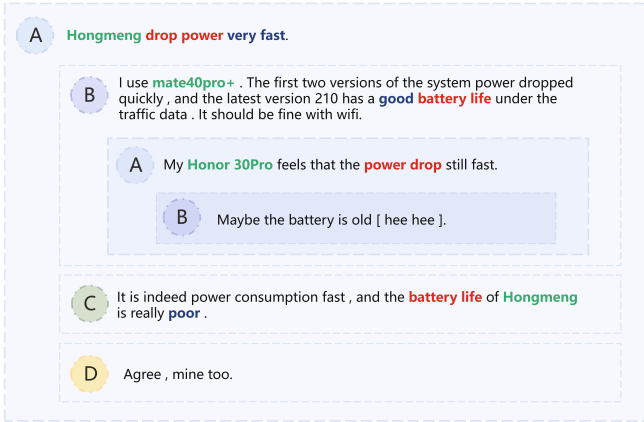


Fig. 1. Illustration of the conversational aspect-based sentiment quadruple analysis. We can extract the following four target-aspect-opinion-sentiment quadruples from this dialog: ('Hongmeng', 'drop power', 'very fast', 'neg'), ('mate40pro+', 'battery life', 'good', 'pos'), ('Honor 30Pro', 'power drop', 'fast', 'neg'), and ('Hongmeng', 'battery life', 'poor', 'neg').

The aforementioned research has been based on short texts such as comments. However, conversational texts are also a significant category in social media, and conducting sentiment analysis on these texts is equally meaningful. To perform fine-grained sentiment analysis on conversational texts, aspect-based sentiment quadruple analysis [12], also known as DiaASQ, has been proposed. As shown in Fig. 1, conversational texts have a natural special structure. Firstly, a conversation consists of multiple participants who may have different stances and views [13]. Secondly, the elements of sentiment quadruples may come from multiple sentences, which we refer to as cross-utterance quadruples. Finally, As the conversation progresses, the topic tends to shift gradually. These characteristics pose new challenges for modeling sentiment analysis on conversational texts.

To address the problem of DiaASQ, Li et al. [12] proposed a new model that used the thread, speaker, and reply views to model the conversation. Their model encodes each utterance separately using pre-trained language models (PLMs) [14], and then models the global information relying on self-attention mechanisms [15] and masking methods. This approach, however, can not fully utilize the powerful contextual modeling capabilities of PLMs, resulting in the loss of some interactive information between adjacent utterance pairs. Therefore, we propose a context-fusion encoding method for DiaASQ, which models the contextual information of the entire thread's speech, rather than model each sentence separately, performing better in extracting cross-utterance quadruples. Meanwhile, we treat extremely short conversations as a whole for context encoding. Furthermore, we incorporate regularized dropout [16] and fast gradient method [17] to improve the robustness of the model.

In summary, the main contributions of this work could be summarized as follows: (1) proposing a context-fusion encoding method that allows the model to better understand context and extract cross-utterance quadruples; (2) incorporating regularized dropout and fast gradient method into the model to enhance its performance; (3) the experimental results have demonstrated that the proposed method achieves an average F1-score of 42.12% in DiaASQ, which is 6.48% higher than the best comparative model, indicating superior performance.

2 Related Work

In this section, we will provide an overview of related work that focuses on sentiment analysis of short texts and the shared task.

2.1 Aspect-Based Sentiment Quadruples Extraction

In the field of aspect-based sentiment analysis (ABSA) for short texts, aspect sentiment quad prediction (ASQP), also referred to as aspect-based sentiment quadruple extraction, has been an active research area [2,3]. Cai et al. [10] were the pioneers to investigate the ABSA quadruple extraction task, with a focus on implicit aspects or opinions. They introduced two new datasets with sentiment quadruple annotations and constructed a series of pipeline baselines by combining existing models to benchmark the task. Zhang et al. [11] proposed a paraphrase modeling strategy to predict sentiment quadruples end-to-end. They transformed the original quadruple prediction task into a text generation problem and solved it using a Seq2Seq modeling paradigm. This approach enabled the full utilization of label semantics, i.e., the meaning of sentiment elements. Later methods have further formalized the task as generating opinion trees [18,19] or structured schema [20].

2.2 Conversational Aspect-Based Sentiment Quadruple Analysis

Conversational aspect-based sentiment quadruple analysis [12] was a new task, and previous work did not consider how to extract sentiment quadruples from conversation text. The shared task provided a model that used a novel labeling scheme based on the grid-tagging method [8], which divided the labeling task into three sub-tasks: detections of entity boundary, entity pair, and sentiment polarity. Compared to pipeline models that required extract-filter-matching processes [10], this approach reduced error propagation and accumulation. Additionally, compared to seq-to-seq approaches [11], it avoided exposure bias. The model first extracted the contextual representation of the sentence through an encoding layer. Then, it proposed a multi-view interaction layer that constructed Thread Mask, Speaker Mask, and Reply Mask, combined with a multi-head self-attention mechanism [15] to strengthen the awareness of the dialogue discourse. Finally, it fused the Rotary Position Embedding (RoPE) [21] and calculated the score between any token pair in terms of the label.

3 Methodology

In this section, we will provide a detailed description of our method. Our model structure is shown in Fig. 2. Overall, we propose a context-fusion encoding method based on the thread and conversation length in the stage of context characterization. We will introduce the adversarial training strategy and regularization technique strategy we used as well.

3.1 Task Introduction

The goal of conversational aspect-based sentiment quadruple analysis is to extract the target-aspect-opinion-sentiment quadruple from conversational texts. The target, aspect, and opinion are continuous words extracted from sentences, and these elements may come from different sentences, referred to as cross-utterance. The sentiment polarity can be classified into three categories: positive, negative, and neutral, based on the extracted three elements. As shown in Fig. 1, a conversation starts from a root post. All subsequent posts are child or grandchild posts of this root post. The so-called thread refers to the subtree derived from the root node of the conversation tree. We treat the root post as a separate thread. Target denotes a particular object(e.g. product or service), while aspect denotes a specific attribute or component of the target. In contrast, category is a broader concept that refers to the class to which the aspect belongs. An opinion term often takes the form of an adjective that conveys the speaker’s evaluation of the aspect. For instance, as shown in Fig. 1, the aspect of “battery life” related to the target “mate40pro+” is mentioned.

Specifically, we represent each dialog as a training sample $D = \{u_1, \dots, u_n\}$ with the corresponding replies $r = \{l_1, \dots, l_n\}$ of utterances, where l_i denotes i^{th} utterance reply to l_i^{th} utterance. To maintain generality, we consider u_1 as root utterance. $t_k = \{u_i, u_{i+1}, \dots, u_j\} (1 \leq i \leq j \leq n)$ represents k -th thread where l_i equal to 1 and $\{l_{i+1} \dots l_j\} \in \{i, i + 1, \dots, j - 1\}$. Each $u_i = \{w_1, \dots, w_{m_i}\}$ denotes i -th utterance text and m_i is the length of utterance of u_i . DiaASQ aims to extract all possible $(target, aspect, opinion, sentiment)$ quadruples, denoted as $Q = \{t, a, o, p\}$ where $\{t, a, o\}$ is the sub-string of dialogue D and $p \in \{pos, neg, other\}$.

3.2 Context Fusion Encoding with Adversarial Training

Thread Fusion. Usually, a dialogue consists of multiple rounds and involves multiple speakers, presenting a complex hierarchical structure. As reported in [12], around 22% of cross-utterance quadruples exist in the Chinese and English datasets. If context encoding is only performed on individual utterance, on the one hand, the outstanding performance of PLMs [14] can not be fully utilized; on the other hand, there is no interaction between different utterances, undoubtedly resulting in the loss of contextual information. Therefore, we propose a contextual fusion method based on thread, which we call “thread fusion”, and use PLMs to better model multiple speakers and different utterances. The method

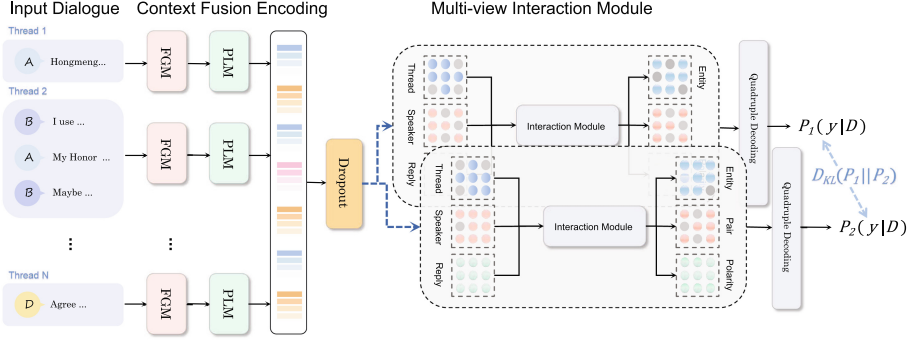


Fig. 2. The overall framework of the proposed method.

merges the utterances in the same thread of conversation into a dialogue segment and treats each segment as a whole for contextual representation encoding.

$$t'_k = \langle [cls], u_i, [sep], u_{i+1}, [sep], \dots, [sep], u_j, [sep] \rangle, \quad (1)$$

$$\mathbf{TH}_k = \mathbf{h}_{cls}, \mathbf{H}_i, \mathbf{h}_{sep}, \dots, \mathbf{h}_{sep}, \mathbf{H}_j, \mathbf{h}_{sep} = \text{PLMs}(t'_k), \quad (2)$$

where u_i, \dots, u_j are the utterances of k -th thread t_k , $[cls]$ and $[sep]$ are the special tokens in PLMs, \mathbf{H}_i and \mathbf{TH}_k means the contextual representation of i -th utterance and k -th thread. We found that the contents discussed in the same thread often have relevance, while the relationships between different threads are relatively weak. This is also the motivation for our proposed thread fusion.

Dialog Fusion. After further analysis of the dataset, it is discovered that some threads in certain conversations are very short in length, containing incomplete quadruples and little information, resulting making no predictions from model. As shown in Table 1, the average length of threads is around 28, with the shortest thread containing only 3 words. Naturally, we consider additional processing for these particularly short threads, by merging them into longer texts. In addition, the maximum length of threads in the Chinese dataset is 257 words, with the longest conversation containing 462 words. It is not applicable to all conversations, as some long conversations may exceed the maximum acceptable length of PLMs. Moreover, long conversations are usually more informative and may introduce noise to the model if merged together.

Taking into account the above two points, we propose treating certain conversations with a length less than a threshold value τ as a whole, and using a PLMs to obtain its global context information. The representation of whole dialog \mathbf{DH} can be constructed as follow:

$$D' = \langle [cls], u_1, [sep], u_2, [sep], \dots, [sep], u_n, [sep] \rangle, \quad (3)$$

$$\mathbf{DH} = \begin{cases} \mathbf{h}_{cls}, \mathbf{H}_1, \mathbf{h}_{sep}, \dots, \mathbf{H}_n, \mathbf{h}_{sep} = \text{PLMs}(D'), & \text{if } \sum_{i=1}^n m_i \leq \tau, \\ \mathbf{TH}_1 || \mathbf{TH}_2 || \dots || \mathbf{TH}_k, & \text{else,} \end{cases} \quad (4)$$

where dialog D' is one training sample connected by $[cls]$ and $[sep]$, m_i is the length of i -th utterance, τ is a controllable hyperparameter that restricts the scope of the processing object, and the operation of “||” is concat.

Adversarial Training. For further improving the performance and robustness of context fusion encoder, we have chosen the Fast Gradient Method (FGM) [17] as our adversarial training technique. FGM is a popular adversarial attack method, which is used in deep learning to generate adversarial examples by perturbing input data to maximize the loss function of the model. It calculates the gradient of the loss function with respect to the input data and perturbs the data in the direction of the gradient with a certain magnitude while maintaining a maximum norm constraint. The perturbations \mathbf{r}_{adv} can be defined as:

$$\mathbf{r}_{adv} = \epsilon \cdot \mathbf{g} / \|\mathbf{g}\|_2 \text{ where } \mathbf{g} = \nabla_s L(D, y), \quad (5)$$

where ϵ is a hyperparameter limiting the size of adversarial perturbations \mathbf{r}_{adv} .

3.3 Quadruple Decoder

Multi-view Interaction. Following Li et al. [12], we construct attention masks M^c and use multi-head self-attention [15] to extract three types of features: dialogue threads, speakers, and reply, where $c \in \{Th, Sp, Rp\}$ and the corresponding values represent thread mask, speak mask and speaker mask, respectively:

$$\mathbf{H}^c = \text{Masked-Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}, \mathbf{M}^c) = \text{Softmax}\left(\frac{(\mathbf{Q}^T \cdot \mathbf{K}) \odot \mathbf{M}^c}{\sqrt{d}}\right) \cdot \mathbf{V}, \quad (6)$$

where $\mathbf{Q} = \mathbf{K} = \mathbf{V} = \mathbf{DH}$ is the representation of whole dialogue. Thread mask $M_{ij}^{Th} = 1$ if the i^{th} and j^{th} token belong to the same dialogue thread; speaker mask $M_{ij}^{Sp} = 1$ if the i^{th} and j^{th} token are derived from the same speaker; and reply mask $M_{ij}^{Rp} = 1$ if the two utterances containing the i^{th} and j^{th} token respectively have a replying relation.

To better guide discourse understanding, the model fuses the Rotary Position Embedding (RoPE) [21] into token representations, which can dynamically encode the relative distance globally between tokens at the dialogue level. And then the score s_{ij}^r indicating the probability of relation label r between w_i and w_j can be calculated as:

$$s_{ij}^r = (\mathbf{R}(\theta, i)\mathbf{v}_i^r)^T(\mathbf{R}(\theta, j)\mathbf{v}_j^r), \quad (7)$$

where $\mathbf{R}(\theta, i)$ is a positioning matrix parameterized by θ and the absolute index i of \mathbf{v}_i^r .

Regularization. Inspired by Liang et al. [16], we improve quadruple decoder using Regularized Dropout (R-Drop), an unsupervised contrastive loss, as the regularization technique. By utilizing the probabilistic nature of the dropout layer, the model’s predictions vary each time. R-Drop passes each training data sample through the model twice, and then uses Kullback-Leibler (KL) divergence to constrain the results of the two predictions, which can be defined by the following formula:

$$\mathcal{L}_{KL} = \frac{\alpha}{2} [\mathcal{D}_{KL}(\mathcal{P}_1^w(y | D) \| \mathcal{P}_2^w(y | D)) + \mathcal{D}_{KL}(\mathcal{P}_2^w(y | D) \| \mathcal{P}_1^w(y | D))], \quad (8)$$

where $\mathcal{P}_1^w(y | D)$ and $\mathcal{P}_2^w(y | D)$ are two distributions of model predictions, α is the coefficient weight to control \mathcal{L}_{KL} .

3.4 Learning

The training loss \mathcal{L}_d of the sum of each subtask can be defined as:

$$\mathcal{L}_k = -\frac{1}{G \cdot N^2} \sum_{g=1}^G \sum_{i=1}^N \sum_{j=1}^N \alpha^k y_{ij}^k \log(p_{ij}^k), \quad (9)$$

$$\mathcal{L}_d = \mathcal{L}_{\text{ent}} + \beta \mathcal{L}_{\text{pair}} + \eta \mathcal{L}_{\text{pol}}, \quad (10)$$

where $k \in \{ \text{ent}, \text{pair}, \text{pol} \}$ indicates a subtask defined by Li [12], N is the total token length in a dialogue, and G is the total training data instances. y_{ij}^k is ground-truth label, p_{ij}^k is the prediction. A tag-wise weighting hyperparameters α^k is applied to counteract the imbalance among label types, where $\alpha^{\text{pair}} = \beta$ and $\alpha^{\text{pol}} = \eta$ are determined by dataset and experimental tuning. The finally loss \mathcal{L} with the loss of R-Drop is:

$$\mathcal{L} = \mathcal{L}_d^1 + \mathcal{L}_d^2 + \mathcal{L}_{KL}, \quad (11)$$

where \mathcal{L}_d^1 and \mathcal{L}_d^2 represent the loss obtained from the model predicting the same sample twice.

4 Experiment

4.1 Datasets and Metrics

Datasets. The corpus consists of posts and comments collected from Weibo, the largest Chinese social media platform. The datasets include both Chinese and English, with the English dataset being translated from the Chinese dataset [12]. As shown in Fig. 1, a dialogue starts from a root post and is composed of replies from multiple speakers. Each reply to the root post is considered as a thread. From a data structure perspective, the multi-thread and multi-turn dialogue forms a tree structure, where each subtree of the root node is a thread. This

data structure provides clear information about the target of each sentence’s reply, which benefits the model’s understanding of context a lot.

The data statistics of datasets are shown in Tables 1. From Table 1, we can see that the English dataset is, on average, slightly longer than the Chinese dataset. The length difference between the shortest and longest samples is very large, regardless of whether it is an utterance, thread, or dialogue.

Metrics. The task of DiaASQ uses exact F1 as the metric, and a sample will be viewed as false unless it matches all four elements exactly. Therefore, the task uses micro F1 and identification F1 [22] respectively for measurements, where micro F1 measures the whole quad, including the sentiment polarity. In contrast, identification F1 does not distinguish the polarity, and is more suitable for evaluating the model’s boundary prediction ability and entity matching ability. Finally, the evaluation criterion for the competition is the average of the four indicators of the Chinese and English datasets.

4.2 Experiment Setting

Due to the similarity in content between the Chinese and English datasets, after initial parameter search, we used the same parameter settings for both datasets. We set the maximum epoch to 30 and trained the model with an early stopping mechanism. The batch size was 1, and evaluation was performed every 100 steps. The initial learning rate was set to 1e-5, and we applied a dropout rate of 0.1 to the intermediate layer. We set the weight α in R-Drop as 1e-4. For the dialogue length threshold τ , we experimented with several different values, including 128, 192, 256, and 512. As shown in Table 1, 512 is already longer than all of dialogues. Following prior work, we used Chinese-Roberta-wwm-base [23] and Roberta-Large [24] as our base encoders for the Chinese and English datasets, respectively.

4.3 Baseline System

We mainly compared some of the latest models of short-text ABSA and dialogue ABSA, as shown below:

- **CRF-Extract-Classify** [10]. A three-stage system (extract, filter, and combine) proposed for the sentence-level quadruple ABSA.
- **SpERT** [25]. A model for joint extraction of entity and relation based on a span-based transformer. The model was slightly modified to support triple-term-extraction and polarity classification.
- **Span-ASTE** [26]. A span-based approach for triplet ABSA extraction. Similarly, it was change to be compatible with the DiaASQ task by editing the last stage of SpanASTE to enumerate triplets.
- **ParaPhrase** [11]. A generative seq-to-seq model for the quadruple ABSA extraction. The model outputs are modified to adapt to DiaASQ task.
- **DiaASQ_{MTV}** [12]. A model to solve the problem of DiaASQ benchmark, which encoding the utterance separately.

Table 1. Statistics on the length of utterance, thread, and dialog in the testset. ‘Utt.’, ‘Thd.’, and ‘Dia.’ respectively refer to utterance, thread, and dialog.

	ZH			EN		
	Utt.	Thd.	Dia.	Utt.	Thd.	Dia.
Avg. Len.	28.88	56.50	218.65	30.56	59.78	231.34
Max. Len.	142	257	462	156	258	481
Min. Len.	3	4	76	3	3	85

Table 2. Performance of the context fusion encoding method in both main experiments and ablation experiments. ‘T-Fusion’ represents the thread fusion method, and ‘D-Fusion₁₂₈’ represents the dialog fusion method with a dialogue length threshold of $\tau=128$.

Model	ZH		EN		Avg.
	Micro	Iden.	Micro	Iden.	
CRF-Extract-Classify	8.81	9.25	11.59	12.80	10.61
SpERT	13.00	14.19	13.07	13.28	13.39
ParaPhrase	23.27	27.98	24.54	26.76	25.64
Span-ASTE	27.42	30.85	26.99	28.34	28.40
DiaASQ _{MTV}	34.94	37.51	33.31	36.80	35.64
Ours	42.79	45.58	38.52	41.60	42.12
w/o FGM, R-Drop	39.96	42.60	35.08	40.13	39.44
w/o D-Fusion ₁₂₈	40.80	44.26	38.20	40.92	41.05
w/o T-Fusion	42.57	44.55	38.51	40.93	41.64

4.4 Results and Analysis

Main Experiment. Table 2 presents the main results of our experiments, demonstrating that our model outperforms all the models with which it is compared. Our best model incorporates thread fusion encoding and dialog fusion encoding with $\tau=128$ and it is trained using FGM and R-Drop. The DiaASQ_{MTV} scores an average of 35.64% on the English and Chinese datasets, while our method exceeds it by approximately 6.48%. This result demonstrates the effectiveness of our approach and theory. Generally, the scores on the Chinese dataset are higher than those on the English dataset.

We also conducted some ablation experiments. First, to verify the effectiveness of context fusion method and eliminate the interference of FGM and R-Drop, we removed these two modules and obtained an average score of 39.44%. Although this score is worse than the main model, it is still 3.8% higher than DiaASQ_{MTV}, further demonstrating that the context fusion method we proposed can help with the context encoding of the model.

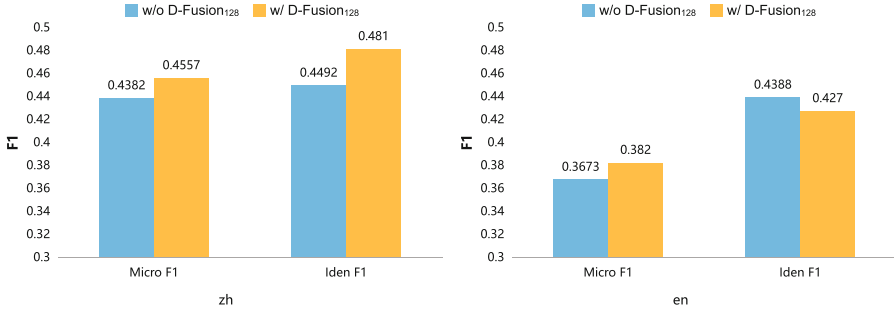


Fig. 3. The comparison of the model with and without D-Fusion₁₂₈ method on the dialog which length less than 128.

In another ablation experiment, we verified whether thread fusion and dialog fusion respectively played a role. The model achieves Average F1 of 41.05% when removed dialog fusion, while achieves 41.64% when removed the thread fusion. We also can find out that dialog fusion had a greater effect than thread fusion. This result was somewhat unexpected, as dialog fusion only processes some short conversations, while thread fusion is effective for all conversations. One possible explanation we propose is that the effect of thread fusion is to improve the accuracy of quadruple extraction within the same thread, whereas many quadruples may not only be cross-utterance but also cross-thread. For cross-thread sentences, dialog fusion can have a greater effect.

In the above experiments, our team achieved the third place in the NLPCC 2023 shared task 4 by obtaining an average score of 41.05% without using the dialog fusion method which is denoted as D-Fusion₁₂₈. In fact, our theoretical best score of 42.12%, which could have achieved a higher ranking, was not submitted due to the competition’s limit of three submission attempts.

4.5 Effectiveness of Dialog Fusion

It is believe that the reason why dialog fusion improves the score is that the model enhances its ability to understand the context of short dialogues. To further verify this, we first identified all sentences with a dialogue length of less than 128 and then compared the model’s F1 scores on these sentences before and after adding D-Fusion₁₂₈. The results are shown in Fig. 3. After adding D-Fusion₁₂₈, the micro F1 and identification F1 scores on the Chinese dataset increased by 1.75% and 3.18% respectively, while the micro F1 score on the English dataset increased by 1.47%. These results support our hypothesis. However, the identification F1 score on the English dataset decreased, indicating that the model’s prediction performance for English boundaries deteriorated after concatenating the dialogues. This may be due to the fact that the English dataset are generally longer than the Chinese dataset (as shown in Table 1) and that English has WordPiece mechanism in PLMs, making the text longer and harder to locate. Overall, this indicates that dialog fusion does improve the accuracy of the model

in understanding and modeling the context of short dialogues. We also experiment different threshold τ for further validation and as the τ increased, there was an overall downward trend in the average score, which is consistent with our hypothesis.

5 Conclusion

This work proposes a context fusion method to enhance the performance of conversational aspect-based sentiment quadruple analysis. Firstly, utterances within the same thread are merged through thread fusion, enabling the model to simultaneously model context information from multiple speakers. Then, dialog fusion is applied to some particularly short dialogues to obtain global information, which effectively improves the model’s performance on shorter dialogues. Through experiments, we conclude that concatenating the entire text of long dialogues actually leads to negative effects. Our model achieved an average F1 score of 42.12%, which is 6.48% higher than the DiaASQ_{MTV}, indicating the effectiveness of our approach.

Acknowledgements. This work was supported by Natural Science Foundation of Guangdong Province (No. 2021A1515011864) and National Natural Science Foundation of China (No. 71472068).

References

1. Phan, H.T., Nguyen, N.T., Hwang, D.: Aspect-level sentiment analysis: a survey of graph convolutional network methods. *Inform. Fusion* **91**, 149–172 (2023)
2. Zhang, W., Li, X., Deng, Y., Bing, L., Lam, W.: A survey on aspect-based sentiment analysis: tasks, methods, and challenges. *CoRR* (2022)
3. Liu, B.: Sentiment analysis and opinion mining. *Synth. Lect. Hum. Lang. Technol.* **5**(1), 1–167 (2012)
4. Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., Hovy, E.: Dual graph convolutional networks for aspect-based sentiment analysis. In: *ACL-IJCNLP* (2021)
5. Zhang, Z., Zhou, Z., Wang, Y.: Ssegcn: syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In: *NAACL-HLT* (2022)
6. Zhou, Y., Liao, L., Gao, Y., Jie, Z., Lu, W.: To be closer: learning to link up aspects with opinions. In: *EMNLP* (2021)
7. Chen, S., Liu, J., Wang, Y., Zhang, W., Chi, Z.: Synchronous double-channel recurrent network for aspect-opinion pair extraction. In: *ACL* (2020)
8. Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., Xia, R.: Grid tagging scheme for end-to-end fine-grained opinion extraction. In: *EMNLP* (2020)
9. Xu, L., Li, H., Lu, W., Bing, L.: Position-aware tagging for aspect sentiment triplet extraction. In: *EMNLP* (2020)
10. Cai, H., Xia, R., Yu, J.: Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In: *ACL-IJCNLP* (2021)
11. Zhang, W., Deng, Y., Li, X., Yuan, Y., Bing, L., Lam, W.: Aspect sentiment quad prediction as paraphrase generation. In: *EMNLP* (2021)

12. Li, B., et al.: Diaasq: A benchmark of conversational aspect-based sentiment quadruple analysis. In: Findings of ACL (2023)
13. Koolagudi, S.G., Rao, K.S.: Emotion recognition from speech: a review. *Inter. J. Speech Technol.* **15**, 99–117 (2012)
14. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K.: Pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT, Bert (2019)
15. Vaswani, A., et al.: Attention is all you need. In: NIPS (2017)
16. Liang, X., et al.: R-drop: regularized dropout for neural networks. In: NeurIPS (2021)
17. Miyato, T., Dai, A.M., Goodfellow, I.: Adversarial training methods for semi-supervised text classification. In: ICLR (2017)
18. Bao, X., Wang, Z., Jiang, X., Xiao, R., Li, S.: Aspect-based sentiment analysis with opinion tree generation. In: IJCAI (2022)
19. Mao, Y., Shen, Y., Yang, J., Zhu, X., Cai, L.: Seq2path: generating sentiment tuples as paths of a tree. In: Findings of ACL (2022)
20. Lu, Y.: Unified structure generation for universal information extraction. In: ACL (2022)
21. Jianlin, S., Yu, L., Pan, S., Murtadha, A., Wen, B., Liu, Y.: Roformer: enhanced transformer with rotary position embedding. CoRR (2021)
22. Barnes, J., Kurtz, R., Oepen, S., Øvrelid, L., Veldal, E.: Structured sentiment analysis as dependency graph parsing. In: ACL/IJCNLP (2021)
23. Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z.: Pre-training with whole word masking for Chinese bert. *IEEE/ACM Trans. Audio Speech Lang. Process.* **29**, 3504–3514 (2021)
24. Liu, Y., et al.: Roberta: a robustly optimized bert pretraining approach. CoRR (2019)
25. Eberts, M., Ulges, A.: Span-based joint entity and relation extraction with transformer pre-training. In: ECAI (2020)
26. Lu, X., Chia, Y.K., Bing, L.: Learning span-level interactions for aspect sentiment triplet extraction In: ACL/IJCNLP (2021)