Inter-sentence Context Modeling and Structure-aware Representation Enhancement for Conversational Sentiment Quadruple Extraction

Yu Zhang[1] **Zhaoman Zhong**[1,2]* **Huihui LV**[1]
[1]School of Computer Science and Engineering, Jiangsu Ocean University, Jiangsu, China
[2]Jiangsu Marine Resources Development Research Institute,
Ministry of Education, Jiangsu, China
zmzhong@jou.edu.cn

Abstract

Conversational aspect-based sentiment quadruple analysis (DiaASQ) is a newly-emergent task aiming to extract quadruples of target-aspect-opinion-sentiment from a conversation text. Existing studies struggle to capture complete dialogue semantics, largely due to inadequate inter-utterance modeling and the underutilization of dialogue structure. To address these issues, we propose an Inter-sentence Context Modeling and Structure-aware Representation Enhancement model (ICMSR) to extract dialogue aspect sentiment quadruple. We design the Dialog Inter-sentence Contextual Enhancer (DICE) module after the sentence-by-sentence encoding phase to enhance inter-sentence interactions and mitigate contextual fragmentation caused by traditional sequential encoding. Moreover, to fully exploit structural information within dialogues, we propose the Dialog Feature Amplifier (DFA), which consists of two submodules: STREAM and SMM. The STREAM module integrates diverse structural dialogue information to generate structure-aware sentence representations, effectively improving the modeling of intra-dialogue structural relations. Furthermore, the Structural Multi-scale Mechanism (SMM) employs a multi-scale modeling approach, simulating varying extents of contextual awareness, thereby enhancing the model's ability to capture cross-sentence structural dependencies. We extensively evaluate our method on benchmark datasets, and the empirical results consistently confirm its effectiveness.

1 Introduction

In the rapidly evolving era of the Internet, vast amounts of user-generated language data are available, making the extraction of hidden user attributes from these statements a crucial and meaningful task in natural language processing (NLP) (Nazir et al., 2020). In real life, a substantial portion of such data exists in the form of conversations, where mining fine-grained knowledge is of significant value across various domains. Conversational aspect-based sentiment quadruple analysis (DiaASQ) is a crucial subtask of aspect-based sentiment analysis (ABSA) (Zhao et al., 2020). The DiaASQ task aims to extract aspect sentiment quadruples from dialogues. Each quadruple consists of four components: the target, the aspect, the opinion, and the associated sentiment polarity, as illustrated in Figure 1.

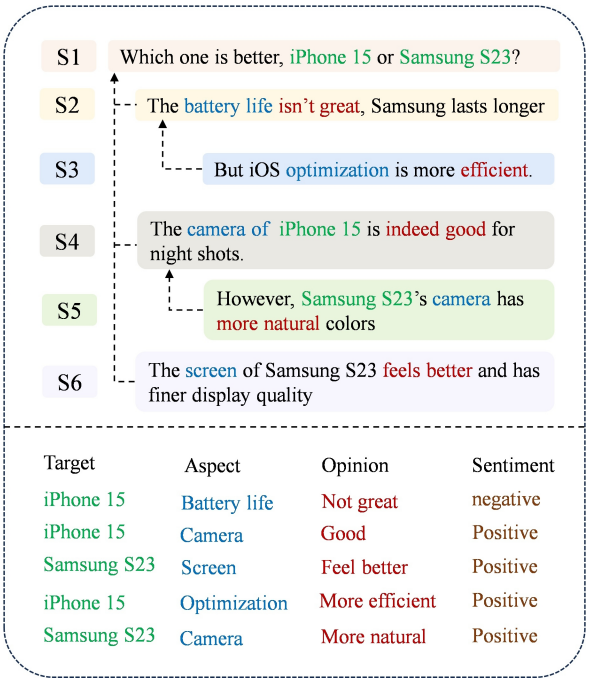


Figure 1: An illustrative example of extracting sentiment quadruples from a dialogue, with target, aspect, opinion, and sentiment components highlighted in distinct colors.

Li et al. (2023) first introduced the DiaASQ task, which aims to extract target entities, aspects, opinions, and their associated sentiment polarities from multi-turn dialogues. Their end-to-end framework utilized max-pooling over dialogue structure fea-

* Corresponding author.

Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pages 17161–17171
November 4-9, 2025

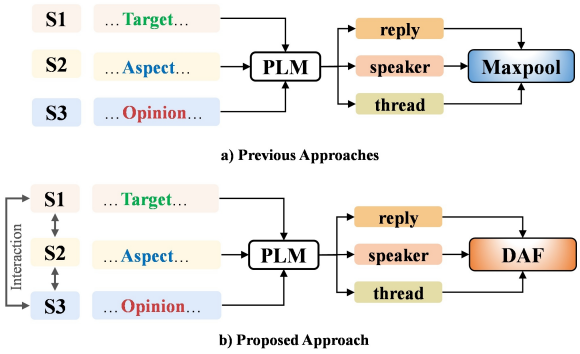


Figure 2: Comparison between previous approaches and our proposed approach.

tures but often failed to capture fine-grained contextual and structural signals. To mitigate this, Jiang et al. (2025) proposed CloBlock, combining pointwise convolution and downsampled self-attention to better model local semantics and global context. Based on generative method, Luo et al. (2024) introduced a segment-assisted denoising and debiasing approach that integrates word-level labeling with utterance-level topic masking to reduce noise.

However, dialog-level quadruple extraction remains inherently challenging due to the fragmented and cross-utterance nature of dialog semantics. As shown in Figure 2, existing approaches face two core limitations. First, they often fail to adequately model inter-utterance dependencies, disrupting contextual coherence and making it difficult to capture long-range semantic information. Second, dialog structural features (speaker, reply, and thread) are underutilized because static pooling-based fusion strategies lack dynamic interactions among these features, resulting in incomplete semantic representations.

To address these issues, we propose ICMSR (Inter-sentence Context Modeling and Structure-aware Representation enhancement), a unified framework inspired by two insights. Initially, inter-sentence modeling can significantly enhance contextual coherence, motivating our DICE module to strengthen the ability of cross-utterance semantic interaction. Subsequently, given that dialog structural features are crucial for capturing discourse dynamics, we propose the DFA module, which integrates a structure-aware fusion layer with a multi-scale refinement mechanism. These components work together to align contextual and structural information, enabling more accurate sentiment quadruple extraction in dialogues.

Experimental results confirm the effectiveness of ICMSR when compared with other DiaASQ approaches that incorporate dialog structural features. Our main contribution are as follows:

- We propose a unified end-to-end framework named ICMSR for Dialogue Sentiment Quadruple Extraction, which jointly tackles two fundamental limitations in existing approaches: inadequate inter-sentence modeling and underutilization of dialogue structural features.
- We design an inter-sentence context modeling module (DICE) to explicitly capture semantic dependencies across utterances, thereby enhancing contextual continuity in multi-turn dialogues.
- We introduce a structure-aware representation enhancement module DFA, which integrates a structural fusion layer (STREAM) and a multi-scale refinement mechanism (SMM) to enrich and strengthen structural representations from multiple perspectives.

2 Related works

2.1 Aspect-Based Sentiment Analysis

Sentiment analysis is a fundamental task in Natural Language Processing (NLP), aimed at identifying emotional tendencies in text. Early studies mainly focused on document-level or sentence-level sentiment classification (Jim et al., 2024; Lin et al., 2022; Zhu et al., 2024), which often overlook nuanced sentiment associated with specific aspects. To address this, Aspect-Based Sentiment Analysis (ABSA) was proposed to capture sentiment polarity toward distinct aspects within a text (Consoli et al., 2022; Huang et al., 2024b; Hellwig et al., 2025). ABSA is typically categorized into extraction-based, classification-based, and hybrid approaches.

Extraction-based methods target sentiment-relevant elements such as aspect terms (AE) (Yang et al., 2021), opinion terms (OE) (Asani et al., 2021), and aspect-opinion pairs (AOE) (Zhao and Yu, 2021). In contrast, classification-based ABSA determines the sentiment polarity of identified aspects (ALSE) (Zeng et al., 2022; Jian et al., 2025). Hybrid methods integrate both extraction and classification, with Aspect-Sentiment Triplet Extraction (ASTE) aiming to jointly identify (Aspect,

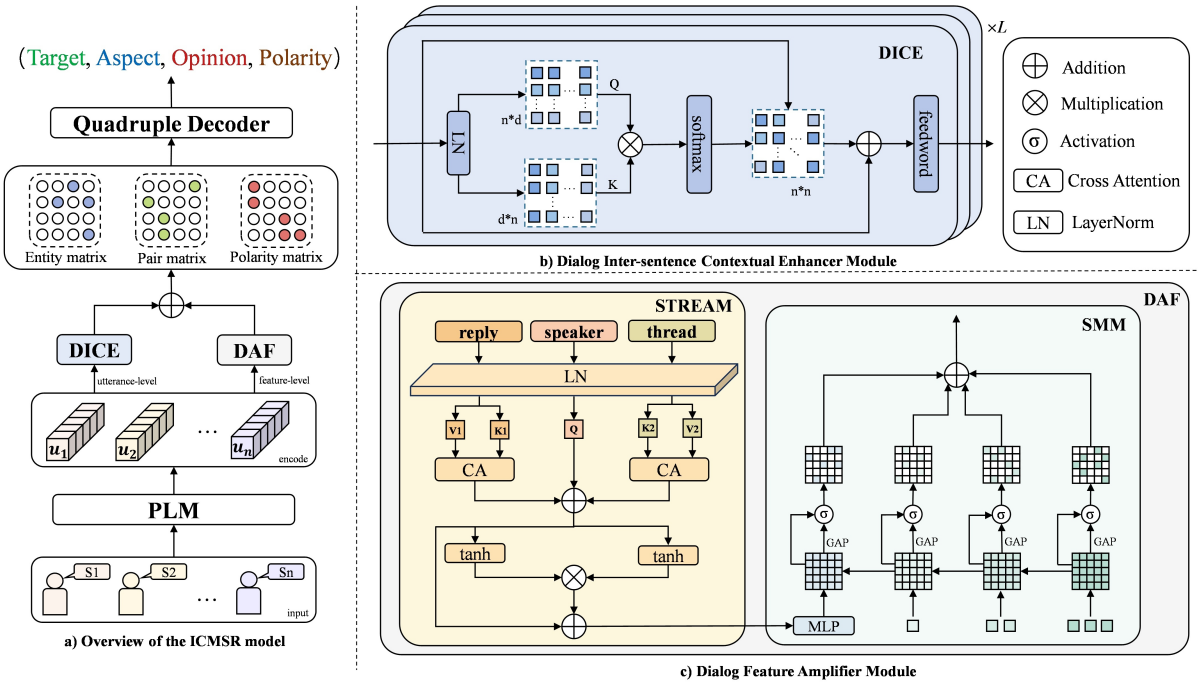


Figure 3: (a) The overview of the proposed ICMSR model. (b) The DICE module adopts a Transformer-like architecture to promote inter-utterance interaction and facilitate long-range dependency modeling in dialogues. (c) The DFA module consists of STREAM and SMM. STREAM employs cross-attention mechanisms to integrate diverse dialogue structural features, while SMM further enhances the fused structural representations through a multi-scale semantic modeling strategy.

Opinion, Sentiment) tuples (Yuan et al., 2023; Jiang et al., 2023; Xu et al., 2025). Further extending this, Aspect-Sentiment Quadruple Extraction (ASQE) incorporates target entities to form (Target, Aspect, Opinion, Sentiment) quadruples for richer sentiment representation (Zhou et al., 2024; Nie et al., 2024; Fu et al., 2024).

Despite advancements, most ABSA methods focus on short, static texts. Dialogue-level sentiment analysis remains underexplored, where modeling the interplay of aspects, opinions, and sentiments across multiple utterances presents new challenges.

2.2 Emotion recognition in conversation

Emotion Recognition in Conversations (ERC) focuses on identifying emotions in individual utterances within multi-turn dialogues, offering a fine-grained understanding of emotional dynamics beyond traditional sentiment analysis (Poria et al., 2019; Shen et al., 2021; Chudasama et al., 2022). It aims to model temporal dependencies across utterances, enabling the detection of both local emotions and their evolution throughout a conversation.

To capture such dynamics, early ERC studies treated dialogue as a sequential process, leveraging Recurrent Neural Networks (RNNs) and their variants. For example, Majumder et al. (Majumder et al., 2019) used RNNs to track speaker states, while Zhang et al. (Zhang et al., 2017) employed LSTMs to model inter-speaker emotional transitions. Jiao et al. (Jiao et al., 2020) introduced a GRU-based framework with dynamic attention and bidirectional encoding to enhance real-time emotion recognition through better context integration.

Beyond sequential models, recent works have explored graph-based methods to represent richer structural dependencies in dialogues. Wang et al. (Wang et al., 2023) proposed a hierarchical extractor using stacked GCN layers to model intra- and inter-speaker dependencies. Zhang et al. (Zhang et al., 2023) utilized dual GATs to jointly encode discourse-level and speaker-level dependencies, capturing complex contextual relations across utterances more effectively.

3 Approach

3.1 Task definition

Given a dialogue $D = \{u_1, u_2, \cdots, u_n\}$ consisting of n utterances, the DiaASQ task focuses on identifying and extracting sentiment quadruples. Aspect-based sentiment quadruples represent targets (t) associated with specific aspects (a)

in the dialogue, along with the expressed opinions (o) and their corresponding sentiment polarity (p). A sentiment analysis quadruple is defined as $Q = \{(t, a, o, p)\}$, where sentiment can be POS (positive), NEG (negative), NEU (neutral), or other categories.

3.2 Textual embedding

We employ a pre-trained language model (PLM) as the encoder to obtain contextualized representations. The PLM operates bidirectionally, enabling it to capture semantic dependencies from both preceding and following tokens within each utterance. To prepare the input, we prepend a [CLS] token and append a [SEP] token to each utterance. The input tokens are first embedded by summing token, segment, and positional embeddings, which are then passed through multiple Transformer layers to model contextual interactions. This yields a sequence of context-aware token representations. The overall encoding process can be summarized as follows:

$$u_i' = \{[CLS], w_1, \cdots, w_k, [SEP]\} \quad (1)$$

$$h_{cls}, h_1, \cdots, h_k, h_{sep} = PLM(u_i') = D' \quad (2)$$

3.3 Dialog Inter-sentence Contextual Enhancer

To address the loss of contextual coherence caused by independent sentence encoding in dialogue quadruple extraction, we propose the Dialog Inter-sentence Contextual Enhancer (DICE). DICE introduces effective cross-sentence interactions and contextual integration after sentence encoding, enhancing the contextual representation of sentences.

DICE initially applies linear transformations to project the input sequence into query and key spaces, thereby explicitly capturing dependencies between sentences:

$$Q_{DICE} = \mathrm{LN}(D') \cdot W_Q^{DICE} \quad (3)$$

$$K_{DICE} = \mathrm{LN}(D') \cdot W_K^{DICE} \quad (4)$$

Here, W_Q^{DICE} and W_V^{DICE} are trainable weight matrices. $\mathrm{LN}(\cdot)$ denotes Layer Normalization, which stabilizes feature distributions before transformation. The inter-sentence similarity is computed through dot product and normalized with softmax to yield the attention score matrix:

$$A_{DICE} = \mathrm{softmax}(\frac{Q_{DICE} \cdot K_{DICE}^\top}{\sqrt{d}}) \quad (5)$$

This matrix $A_{DICE} \in \mathbb{R}^{n \times n}$ captures the strength of dependencies between sentences in the dialogue. Based on this attention matrix, DICE further aggregates sentence representations in a weighted manner to integrate contextual information from other sentences, while employing a residual connection and a learnable scaling factor γ to dynamically control the degree of fusion:

$$H = A_{DICE} D' \quad (6)$$

$$X = \gamma H + D' \quad (7)$$

The scaling factor γ is initialized to zero, so the model initially depends on original sentence representations and gradually incorporates contextual features as training advances, ensuring stable feature enhancement. After cross-sentence interaction modeling, DICE uses a feed-forward network to nonlinearly transform sentence representations for better expressiveness:

$$X' = \mathrm{ReLU}(XW_1 + b_1)W_2 + b_2 \quad (8)$$

In this formulation, b_1 and b_2 denote bias terms. To capture more complex dialogue structures and inter-sentence dependencies, DICE is applied in a multi-layered manner, where each layer refines sentence representations based on contextual signals from other utterances. Formally, the representations at the l-th layer are computed as:

$$D^{(l)} = DICE(D^{(l-1)}) \quad (9)$$

Specifically, $D^{(0)} = D'$ denotes the initial sentence-level representations. After stacking L layers, the final enhanced representations are obtained as:

$$\hat{D}_1 = D^{(L)} \quad (10)$$

3.4 Dialog Structural Feature Amplifier

The DFA module aims to strengthen the model's capability to capture cross-utterance dependencies by leveraging dialogue structural features. It is composed of two submodules: STREAM, which aggregates heterogeneous structural information, and SMM, which enhances the model's perception of cross-utterance contextual relationships.

3.4.1 Speaker-Thread-Reply Enhanced Aggregation Module

To effectively leverage structural information within multi-turn dialogues, we propose a Speaker-Thread-Reply Enhanced Aggregation Module (STREAM), which dynamically fuses speaker roles, topic threads, and reply dependencies.

STREAM first applies layer normalization to the input speaker, thread, and reply features, followed by linear projections to obtain their corresponding query, key, and value representations:

$$Q = W_q \cdot LayerNorm(speak) \quad (11)$$

$$K_{thr}, V_{thr} = W_{thr} \cdot LayerNorm(thread) \quad (12)$$

$$K_{rep}, V_{rep} = W_{rep} \cdot LayerNorm(reply) \quad (13)$$

Where W_q, W_{thr} and W_{rep} denote trainable projection matrices. This transformation ensures that heterogeneous structural features are mapped into a shared representation space, facilitating subsequent interactions. Next, STREAM calculates two independent attention distributions to capture the dependencies between speaker features and thread or reply features, respectively:

$$Attn_1 = \mathrm{Softmax}(\frac{Q \cdot K_{thr}^T}{\sqrt{d}} \cdot \tau) \cdot V_{thr} \quad (14)$$

$$Attn_2 = \mathrm{Softmax}(\frac{Q \cdot K_{rep}^T}{\sqrt{d}} \cdot \tau) \cdot V_{rep} \quad (15)$$

Here, τ is a temperature factor to adjust attention sharpness. Thread attention aligns target and aspect information, while reply attention captures opinion-sentiment interactions. The outputs are then fused with learnable weights to dynamically balance the contributions of each structure:

$$s_{fusion} = \alpha_{sp} \cdot spreak + \alpha_{thr} \cdot Attn_1 + \alpha_{rep} \cdot Attn_2 \quad (16)$$

The parameters α_{sp}, α_{thr} and α_{rep} are trainable to control the fusion degree. To further enhance the fused representation, STREAM incorporates an Improved Enhancement Layer (IEL). IEL projects the fused features to a higher-dimensional space and splits them into two parallel branches. And each branch undergoes non-linear transformation with residual connection:

$$H_1, H_2 = Chunk(s_{fusion}) \quad (17)$$

$$H_1' = \tanh(H_1 \cdot W_1) \quad (18)$$

$$H_2' = \tanh(H_2 \cdot W_2) \quad (19)$$

Finally, the two branches are combined via element-wise multiplication and projected back. The final output adds a residual connection to the original speaker features:

$$F_{STREAM} = (H_1' \odot H_2') \cdot W_{out} + speak \quad (20)$$

3.4.2 Structural Multi-scale Mechanism

To further enhance the structural representations obtained from STREAM, we design the Structural Multi-Scale Mechanism (SMM). SMM aims to capture hierarchical dependencies and multi-scale semantic patterns within the aggregated dialogue structure features, improving the model's capability to handle varying granularities of structural information.

SMM begins by processing the input features through four parallel layers with different expansion rates to simulate multi-scale perception:

$$y_0 = \mathrm{ReLU}((\mathrm{MLP}(x) \cdot W_0) \quad (21)$$

$$y_i = \mathrm{ReLU}(\mathrm{ReLU}(y_{i-1} \cdot W_i^{(1)}) \cdot W_i^{(2)}) \quad (22)$$

Where $W_i^{(1)}$ and $W_i^{(2)}$ are projection matrices corresponding to the i-th scale layer, where different expansion rates allow each branch to model structure at varying resolutions. Next, SMM performs global average pooling to summarize each scale representation and computes scale-specific importance weights:

$$\tilde{w_i} = \sigma(W \cdot (GAP(y_i)) + b) \quad (23)$$

These weights are normalized via softmax to emphasize the most informative scales:

$$[w_0, w_1, w_2, w_3] = \mathrm{softmax}([\tilde{w_0}, \tilde{w_1}, \tilde{w_2}, \tilde{w_3}]) \quad (24)$$

Finally, SMM aggregates the multi-scale features using the learned weights:

$$F_{SMM} = w_0 \cdot y_0 + w_1 \cdot y_1 + w_2 \cdot y_2 + w_3 \cdot y_3 \quad (25)$$

$$\hat{D}_2 = F_{STREAM} + F_{SMM} \quad (26)$$

In summary, STREAM captures cross-structural dependencies, while SMM further enriches these representations at multiple semantic scales. Together, they form a comprehensive structural encoder for accurate dialogue quadruple extraction.

3.5 Decoding and learning

Given the multi-task nature of the DiaASQ task, we adopt a tag-wise representation approach for modeling. Based on this, we compute the association score between token pairs concerning a relation label r.

$$u_i^r = MLP^r(d_i) \quad (27)$$

$$s_{ij}^r = (u_i^r)^T u_j^r \quad (28)$$

$$p_{ij}^k = \mathrm{softmax}((s_{ij}^{\in k}; s_{ij}^{k_1}; \cdots; s_{ij}^{k_n})) \quad (29)$$

Table 1: Main results on the DiaASQ dataset. The best performance for each metric is highlighted in **bold**, and the second best performance is underlined. 'T/A/O' denote Target, Aspect, and Opinion respectively. All baseline results are cited from the original papers.

Model	ZH								EN							
	T	A	O	T-A	T-O	A-O	Mi-F1	Id-F1	T	A	O	T-A	T-O	A-O	Mi-F1	Id-F1
ChatGPT(0-shot)	/	/	/	23.86	10.55	15.82	13.77	18.15	/	/	/	23.26	16.07	14.34	10.98	12.99
ChatGPT(1-shot)	/	/	/	29.90	17.48	25.59	18.26	20.56	/	/	/	26.18	20.20	21.20	13.20	14.67
ChatGPT(5-shot)	68.78	57.87	36.45	34.98	42.48	27.43	18.41	20.59	68.05	53.22	45.08	28.76	37.24	25.36	15.26	17.17
ParaPhrase	/	/	/	37.81	34.32	27.76	23.27	27.98	/	/	/	37.22	32.19	30.78	24.54	26.76
Span-ASTE	/	/	/	44.13	34.46	32.21	27.42	30.85	/	/	/	42.19	30.44	45.90	26.99	28.34
DiaASQ	90.23	76.94	59.35	48.61	43.31	45.44	34.94	37.51	88.62	74.71	60.22	47.91	45.58	44.27	33.31	36.80
Overall-QPN	/	/	/	52.86	50.98	53.33	37.77	43.56	/	/	/	50.70	49.46	50.31	35.37	39.76
IFusionQuad	91.69	75.90	60.96	54.68	51.81	50.04	41.53	44.56	88.31	74.23	63.48	52.65	51.82	51.94	35.96	41.49
SARA	**92.60**	77.30	61.62	**56.88**	51.65	**54.77**	42.51	**45.75**	88.88	74.98	**64.85**	**54.64**	51.82	**54.30**	**39.40**	42.64
ICMSR	91.39	**78.01**	**63.30**	56.69	**52.53**	54.59	**42.55**	45.20	**88.91**	**75.04**	63.93	54.23	**52.67**	51.61	39.36	**44.06**

where r represents the type of relationship between two tokens. Thus, s_{ij}^{r} represents the likelihood score that a token pair shares a specific relation. Then a fully connected layer is applied to compute the probability distribution of the relation matrix.

To minimize training loss during the training process, cross-entropy is used to compute the loss for each task, and the final loss is obtained by aggregating these individual losses:

$$L_c = -\frac{1}{G \cdot N^2} \sum_{g=1}^{G} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha^c y_{ij}^c \log(p_{ij}^c) \quad (30)$$

$$L = L_{ent} + L_{pair} + L_{pol} \quad (31)$$

where $c \in \{ent, pair, pol\}$ is the type of the subtask, G is training instances, N is the length of a total dialogue. y_{ij}^c represents ground-truth label and p_{ij}^c is the prediction.

4 Experiment and discussion

4.1 Experiment settings

4.1.1 Datasets and metrics

To effectively address the DiaASQ task, we selected the dataset constructed by Li et al. as the foundation for our study. This dataset is sourced from Weibo, a social media platform where the data typically exhibits multi-turn interactions and dynamic changes, making it highly suitable for DiaASQ task research. Specifically, the Chinese version of the dataset contains 7452 discourses and 5742 quaternions, while the English version includes 5514 quaternions.

Based on previous research, we used F1 as evaluation metrics in the experiment. A quaternion is considered a successful prediction when the predicted target entity, aspect entity, opinion entity, and sentiment polarity match the corresponding ground truth entities and polarity.

4.1.2 Implementation Details

The proposed ICMSR model is implemented using the PyTorch framework and trained on an NVIDIA RTX 4090 GPU. Its parameter scale is 124M. We adopt PLMs as encoders, specifically using chinese-roberta-wwm-ext[1] for Chinese and roberta-large[2] for English, following common practices in prior work (Li et al., 2023). The main model is trained using the Adam optimizer, with a learning rate of 1e-3 to ensure numerical stability. We set the batch size to 2 and train the model for 10 epochs, with each epoch requiring about 1min54s. These hyperparameter settings are selected based on preliminary experiments (Jiang et al., 2025) on the development set and have been found to yield stable training and competitive performance across datasets.

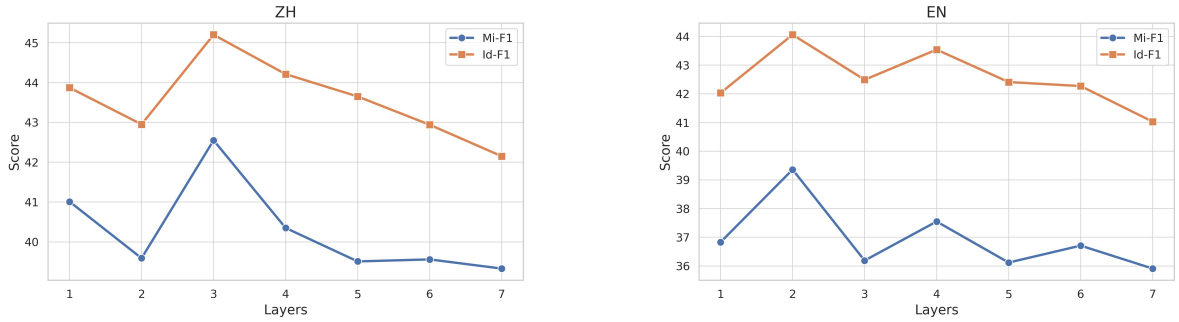
4.2 Baselines

We compare our model against a range of representative baselines with utilizing dialog structural features. ChatGPT-based approaches use ChatGPT-3.5-turbo[3]. Related experimental results are cited from (Huang et al., 2024a) and (Zhou et al., 2024). Span-based methods such as Span-ASTE (Xu et al., 2021) jointly extract entities and relations or triplets via Transformer encoders, with strategies to model overlapping spans and maintain sentiment consistency. Generative approaches like ParaPhrase (Zhang et al., 2021) treat the task as sequence generation, leveraging paraphrasing for unified representation learning. Dialogue-specific frameworks include DiaASQ (Li et al., 2023), which incorporates multi-view interaction and distance-aware tagging, SARA (Liu et al., 2025), which combines

[1] https://huggingface.co/hfl/chinese-roberta-wwm-ext

[2] https://huggingface.co/FacebookAI/roberta-large

[3] https://platform.openai.com



(a) ZH (b) EN

Figure 4: Performance impact of the number of DICE layers

with a relation-augmented grid tagging scheme, and IFusionQuad (Jiang et al., 2025), a hybrid system combining CloBlock and Biaffine attention for fine-grained semantic fusion.

4.3 Main results

We compare ICMSR with several strong baselines on both Chinese (ZH) and English (EN) datasets across three subtasks: entity recognition, pairwise relation extraction (T-A, T-O, A-O), and sentiment quadruple extraction. As shown in Table 1, ICMSR consistently achieves the best performance. On ZH, it surpasses IFusionQuad with accuracy gains of 2.01, 0.72, and 4.55 on the three pairwise subtasks, and improves Mi-F1 and Id-F1 by 1.02 and 0.64, respectively, on the full task. On EN, it achieves gains of 3.40 Mi-F1 and 2.57 Id-F1, mainly due to improved recall. Compared with DiaASQ, ICMSR brings larger improvements: +7.61/+7.69 on ZH and +6.05/+7.26 on EN. It is worth noting that although ChatGPT does not perform well on the complex task of dialogue-based sentiment quadruple extraction, it struggles to capture the relationships among different sentiment elements in multi-turn dialogues even under few-shot prompting. These results confirm the effectiveness of both DICE and DFA in capturing cross-utterance dependencies and leveraging dialogue structure.

4.4 Ablation study

To examine the contributions of each module in ICMSR, we perform ablation studies on both ZH and EN datasets by removing DICE, DFA, and its subcomponents STREAM and SMM. As shown in Table 2, the full model achieves the best performance across all settings. Excluding DICE results in a noticeable drop in Mi-F1, especially on ZH, highlighting the importance of inter-sentence context modeling. Further removing DFA leads to a more substantial decline, particularly on EN, demonstrating the critical role of structural information. Among DFA's submodules, STREAM contributes more prominently on EN, indicating that speaker and reply-aware features help model long-range dialogue dependencies. In contrast, removing SMM causes a larger performance drop on ZH, suggesting that multi-scale structural enhancement is particularly beneficial for Chinese dialogues with more complex discourse organization.

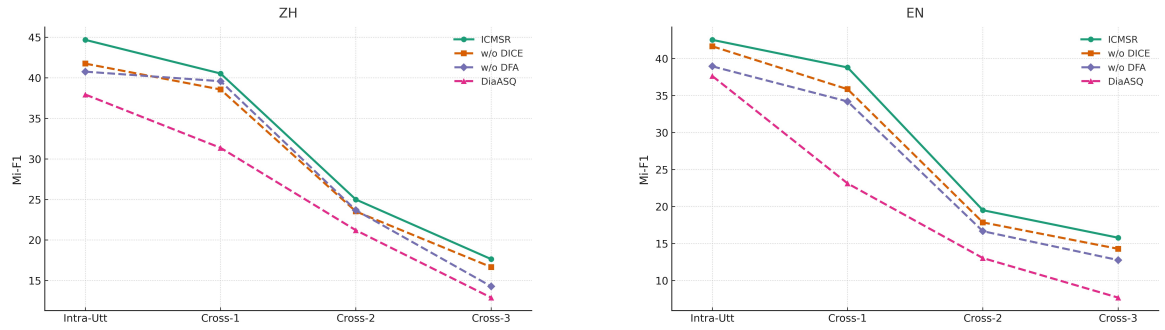
Table 2: Ablation study on the DiaASQ dataset. Mi-F1 and Id-F1 scores are reported on Chinese (ZH) and English (EN) subsets by removing each module from ICMSR.

Model	ZH		EN	
	Mi-F1	Id-F1	Mi-F1	Id-F1
w/o DICE	40.62	44.49	37.97	43.48
w/o DFA	39.30	42.98	37.53	41.65
w/o STREAM	39.47	43.61	36.75	41.00
w/o SMM	39.34	42.65	38.10	42.86
ICMSR	42.55	45.20	39.36	44.06

4.5 The details of model

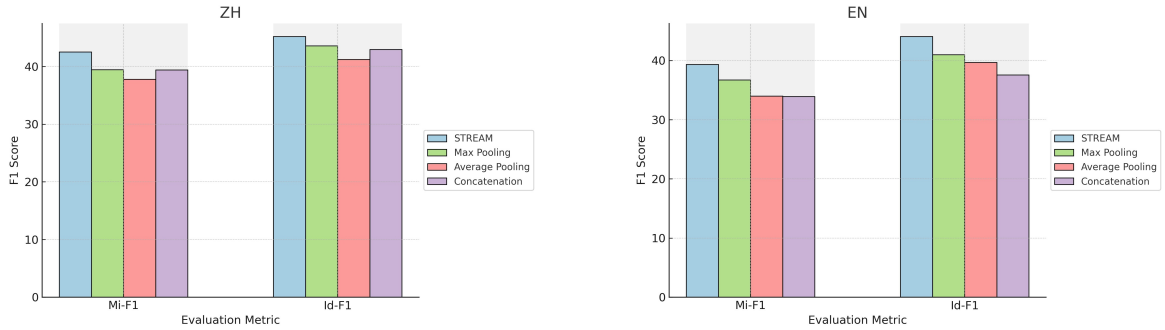
4.5.1 Effects of DICE layer depth

We investigate how varying the number of DICE layers (1–7) affects model performance on ZH and EN datasets (Figure 4). Results exhibit a non-monotonic trend, peaking at moderate depths—3 layers for ZH (42.55 Mi-F1, 45.20 Id-F1) and 2 layers for EN (39.36 Mi-F1, 44.06 Id-F1)—with performance declining at greater depths due to increased noise or redundancy. The sharper decline on ZH (Mi-F1 dropping to 39.33 at 7 layers) suggests Chinese dialogues' higher discourse complexity makes deeper stacking counterproductive,



(a) ZH (b) EN

Figure 5: Performance under different levels of cross-utterance



(a) ZH (b) EN

Figure 6: Effects of different feature fusion methods on model performance

whereas flatter English dialogue structures are less depth-sensitive.

4.5.2 Effects of cross utterance

We conduct a detailed analysis of cross-utterance quadruple extraction to assess the model's ability to capture long-range dependencies. As illustrated in Figure 5, performance across all models declines as the utterance distance increases. Notably, at the Cross-3 level, baseline systems such as DiaASQ fail to extract valid quadruples, indicating their limitations in handling long-span dependencies.

In contrast, our full model ICMSR consistently outperforms baselines across all levels, maintaining strong performance even in highly fragmented contexts. This highlights its effectiveness in modeling inter-sentence interactions and leveraging structural cues. The clear performance drop in ablated variants (w/o DICE and w/o DFA) further validates the complementary contributions of both contextual modeling and structure-aware representation in tackling cross-utterance challenges.

4.5.3 Effects of different fusion methods

To examine the effectiveness of different fusion strategies for dialog structural features, we conducted comparative experiments on ZH and EN datasets. As shown in Figure 6, STREAM consistently achieves the highest Mi-F1 and Id-F1 scores, significantly outperforming traditional methods (Max Pooling, Average Pooling, and Concatenation), particularly with an improvement of over 3 points in Id-F1 on EN. This indicates STREAM's superior ability in capturing inter-utterance structural dependencies, demonstrating robustness and generalizability across languages.

5 Conclusion

We propose a model named Inter-sentence Context Modeling and Structure-aware Representation Enhancement for extracting sentiment quadruples in aspect-based sentiment analysis. Compared with previous approaches, our model improves the capacity for cross-utterance interaction, enabling more comprehensive modeling of contextual dependencies within dialogues. Furthermore, it demon-

strates strong performance in leveraging dialogue structural features. This is attributed to the introduction of a Dialogue Feature Amplifier module, which precisely models the interactions among various structural elements and enhances their representations across multiple semantic scales. Extensive experiments verify the effectiveness of each component in the model. Overall, our approach surpasses strong baseline models and achieves well performance.

Limitations

1. Limited Cross-Domain Validation: Currently, the DiaASQ dataset is the only available benchmark for dialogue aspect-based sentiment quadruple analysis. Due to its specific domain characteristics and dialogue structures, the generalization capability of our model to other domains or platforms remains unverified. Thus, further exploration and validation across diverse domains are essential to assess and improve the model's cross-domain adaptability.

2. Dependency on Pre-trained Language Models: Our approach relies heavily on different pre-trained language models for encoding context across languages. Although these models perform effectively on standard benchmark datasets, their performance may significantly degrade in real-world scenarios characterized by substantial data distribution shifts or limited computational resources. Future studies will focus on enhancing cross-lingual generalizability and exploring robust solutions under resource-constrained conditions.

3. Insufficient Exploration of Multi-party Dialogue Scenarios: Despite our model achieving superior performance compared to baseline models in capturing multi-turn, cross-utterance interactions, its effectiveness remains limited in more complex multi-party conversational settings. The intricate interaction patterns present in multi-party dialogues pose significant challenges that are not adequately addressed in the current approach. In future work, we plan to incorporate additional features, such as syntactic information, and analyze them in conjunction with dialog structural features to further investigate their impact on this task.

Ethics Statement

This work focuses on dialog-level sentiment quadruple extraction. All datasets used are publicly available from prior research, without involving any new data collection from human participants, and we strictly adhere to the original licenses and usage guidelines. Before conducting experiments, we manually verified that the datasets had been anonymized and contained no personally identifiable information.

Throughout the research process, we carefully referred to ethical sheet (Mohammad, 2022) on automatic emotion recognition and designed our experiments and data usage with the aim of minimizing potential risks. In line with broader ethical discussions on dialog analysis (Ruane et al., 2019; Luxton, 2020), we emphasize that deploying dialog-based sentiment analysis systems in public domains and cross-cultural contexts requires particular attention to transparency, informed consent, and the risk of misuse. Given that dialog-level sentiment analysis is still at an exploratory stage, the technology is not yet ready for real-world deployment. Accordingly, both existing studies and the present work remain limited to academic inquiry, aiming to advance a fine-grained understanding of dialog sentiment within natural language processing. Looking forward, we advocate for the introduction of appropriate ethical and regulatory mechanisms in future practical applications to ensure that such technologies are adopted in a socially responsible and ethically sound manner.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (72174079), the Jiangsu "Qinglan Project" Ex-cellent Teaching Team in Big Data (2022-29), and the Lianyungang Key R&D Program (Industry Foresight and Key Core Technologies) Project (CG2323).

References

Elham Asani, Hamed Vahdat-Nejad, and Javad Sadri. 2021. Restaurant recommender system based on sentiment analysis. *Machine Learning with Applications*, 6:100114.

Vishal Chudasama, Purbayan Kar, Ashish Gudmalwar, Nirmesh Shah, Pankaj Wasnik, and Naoyuki Onoe. 2022. M2fnet: Multi-modal fusion network for emotion recognition in conversation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4652–4661.

Sergio Consoli, Luca Barbaglia, and Sebastiano Manzan. 2022. Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. *Knowledge-Based Systems*, 247:108781.

Yiheng Fu, Xiaoliang Chen, Duoqian Miao, Xiaolin Qin, Peng Lu, and Xianyong Li. 2024. Label-semantics enhanced multi-layer heterogeneous graph convolutional network for aspect sentiment quadruplet extraction. *Expert Systems with Applications*, 255:124523.

Nils Constantin Hellwig, Jakob Fehle, and Christian Wolff. 2025. Exploring large language models for the generation of synthetic training samples for aspect-based sentiment analysis in low resource settings. *Expert Systems with Applications*, 261:125514.

Peijie Huang, Xisheng Xiao, Yuhong Xu, and Jiawei Chen. 2024a. Dmin: A discourse-specific multi-granularity integration network for conversational aspect-based sentiment quadruple analysis. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 16326–16338.

Xiaosai Huang, Jing Li, Jia Wu, Jun Chang, Donghua Liu, and Kai Zhu. 2024b. Flexibly utilizing syntactic knowledge in aspect-based sentiment analysis. *Information Processing & Management*, 61(3):103630.

Zhongquan Jian, Daihang Wu, Shaopan Wang, Yancheng Wang, Junfeng Yao, Meihong Wang, and Qingqiang Wu. 2025. Agcl: Aspect graph construction and learning for aspect-level sentiment classification. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 841–854.

Baoxing Jiang, Shehui Liang, Peiyu Liu, Kaifang Dong, and Hongye Li. 2023. A semantically enhanced dual encoder for aspect sentiment triplet extraction. *Neurocomputing*, 562:126917.

Haoyu Jiang, Xiaoliang Chen, Duoqian Miao, Hongyun Zhang, Xiaolin Qin, Xu Gu, and Peng Lu. 2025. Ifusionquad: A novel framework for improved aspect-based sentiment quadruple analysis in dialogue contexts with advanced feature integration and contextual cloblock. *Expert Systems with Applications*, 261:125556.

Wenxiang Jiao, Michael Lyu, and Irwin King. 2020. Real-time emotion recognition via attention gated hierarchical memory network. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 8002–8009.

Jamin Rahman Jim, Md Apon Riaz Talukder, Partha Malakar, Md Mohsin Kabir, Kamruddin Nur, and Mohammed Firoz Mridha. 2024. Recent advancements and challenges of nlp-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, page 100059.

Bobo Li, Hao Fei, Fei Li, Yuhan Wu, Jinsong Zhang, Shengqiong Wu, Jingye Li, Yijiang Liu, Lizi Liao, Tat-Seng Chua, and Donghong Ji. 2023. DiaASQ: A benchmark of conversational aspect-based sentiment quadruple analysis. In *Findings of ACL*, pages 13449–13467.

Szu-Yin Lin, Yun-Ching Kung, and Fang-Yie Leu. 2022. Predictive intelligence in harmful news identification by bert-based ensemble learning model with text sentiment analysis. *Information Processing & Management*, 59(2):102872.

Xiaoyong Liu, Miao Hu, Chunlin Xu, and Zhiguo Du. 2025. Sara: Span-aware framework with relation-augmented grid tagging for conversational aspect-based sentiment quadruple analysis. *Engineering Applications of Artificial Intelligence*, 154:110915.

Xianlong Luo, Meng Yang, and Yihao Wang. 2024. Overcome noise and bias: Segmentation-aided multi-granularity denoising and debiasing for enhanced quarduples extraction in dialogue. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 839–856.

David D Luxton. 2020. Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, 98(4):285.

Navonil Majumder, Soujanya Poria, Devamanyu Hazarika, Rada Mihalcea, Alexander Gelbukh, and Erik Cambria. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6818–6825.

Saif M Mohammad. 2022. Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, 48(2):239–278.

Ambreen Nazir, Yuan Rao, Lianwei Wu, and Ling Sun. 2020. Issues and challenges of aspect-based sentiment analysis: A comprehensive survey. *IEEE Transactions on Affective Computing*, 13(2):845–863.

Yu Nie, Jianming Fu, Yilai Zhang, and Chao Li. 2024. Modeling implicit variable and latent structure for aspect-based sentiment quadruple extraction. *Neurocomputing*, 586:127642.

Soujanya Poria, Navonil Majumder, Rada Mihalcea, and Eduard Hovy. 2019. Emotion recognition in conversation: Research challenges, datasets, and recent advances. *IEEE access*, 7:100943–100953.

Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational ai: Social and ethical considerations. *AICS*, 2563:104–115.

Weizhou Shen, Junqing Chen, Xiaojun Quan, and Zhixian Xie. 2021. Dialogxl: All-in-one xlnet for multi-party conversation emotion recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 13789–13797.

Binqiang Wang, Gang Dong, Yaqian Zhao, Rengang Li, Qichun Cao, Kekun Hu, and Dongdong Jiang. 2023. Hierarchically stacked graph convolution for emotion recognition in conversation. *Knowledge-Based Systems*, 263:110285.

Guangtao Xu, Zhihao Yang, Bo Xu, Ling Luo, and Hongfei Lin. 2025. Span-based syntactic feature fusion for aspect sentiment triplet extraction. *Information Fusion*, page 103078.

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. Association for Computational Linguistics.

Heng Yang, Biqing Zeng, Jianhao Yang, Youwei Song, and Ruyang Xu. 2021. A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction. *Neurocomputing*, 419:344–356.

Li Yuan, Jin Wang, Liang-Chih Yu, and Xuejie Zhang. 2023. Encoding syntactic information into transformers for aspect-based sentiment triplet extraction. *IEEE Transactions on Affective Computing*, 15(2):722–735.

Jiandian Zeng, Tianyi Liu, Weijia Jia, and Jiantao Zhou. 2022. Relation construction for aspect-level sentiment classification. *Information Sciences*, 586:209–223.

Duzhen Zhang, Feilong Chen, and Xiuyi Chen. 2023. Dualgats: Dual graph attention networks for emotion recognition in conversations. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7395–7408.

Ruo Zhang, Atsushi Ando, Satoshi Kobashikawa, and Yushi Aono. 2017. Interaction and transition model for speech emotion recognition in dialogue. In *INTERSPEECH*, pages 1094–1097.

Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. 2021. Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219. Association for Computational Linguistics.

Anping Zhao and Yu Yu. 2021. Knowledge-enabled bert for aspect-based sentiment analysis. *Knowledge-Based Systems*, 227:107220.

He Zhao, Longtao Huang, Rong Zhang, Quan Lu, and 1 others. 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.

Changzhi Zhou, Zhijing Wu, Dandan Song, Linmei Hu, Yuhang Tian, and Jing Xu. 2024. Span-pair interaction and tagging for dialogue-level aspect-based sentiment quadruple analysis. In *Proceedings of the ACM Web Conference 2024*, pages 3995–4005.

Wenhao Zhu, Jiayue Qiu, Ziyue Yu, and Wuman Luo. 2024. A survey on personalized document-level sentiment analysis. *Neurocomputing*, 609:128449.