TCDA: Thread-Constrained Discourse-Aware Modeling for Conversational Sentiment Quadruple Analysis

Xinran Li[1] , **Xinze Che**[1] , **Yifan Lyu**[1] , **Zhiqi Huang**[2] and **Xiujuan Xu**[1,*]
[1]School of Software, Dalian University of Technology, Dalian, China
[2]School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
963707605@mail.dlut.edu.cn , xjxu@dlut.edu.cn

Abstract

Conversational Aspect-based Sentiment Quadruple Analysis (DiaASQ) needs to capture the complex interrelationships in multiple rounds of dialogues. Existing methods usually employ simple Graph Convolutional Networks (GCN), which introduce structural noise and fail to consider the temporal sequence of the dialogues, or use standard RoPE, which implicitly captures relative distances in a flat sequence but cannot clearly separate the token-level syntactic order from the utterance-level progression, and may suffer from the Distance Dilution problem. To address these issues, we propose a new framework that combines Thread-Constrained Directed Acyclic Graph (TC-DAG) and Discourse-Aware Rotary Position Embedding (D-RoPE). Specifically, TC-DAG filters out cross-thread noise based on thread constraints, maintains global connectivity through root anchoring, and incorporates the temporal sequence of the dialogues. D-RoPE aligns multi-layer semantics using dual-stream projection and multi-scale frequency signals, captures thread dependencies using tree-like distances, and alleviates the token-level Distance Dilution problem by incorporating utterance-level progressions. Experimental results on two benchmark datasets demonstrate that our framework achieves state-of-the-art performance.

1 Introduction

With the rapid proliferation of online social media and real-time communication platforms, the task of Conversational Aspect-based Sentiment Quadruple Analysis (*DiaASQ*) [Li *et al.*, 2023] has emerged to meet the growing demand for fine-grained sentiment understanding in conversations. As shown in Figure 1 and Table 1, the goal of *DiaASQ* is to automatically extract all existing sentiment quadruples (t, a, o, s) from the given multi-round conversation. In this formulation, **target** t (the object of discussion), **aspect** a (the specific attribute of the target) and **opinion** o (the subjective expression about the aspect) correspond to specific text spans in the conversation. Meanwhile, **sentiment** s represents the emotional polarity, which is usually classified as positive, negative or neutral. Different from traditional sentence-level sentiment analysis [Zhang *et al.*, 2021; Mao *et al.*, 2022], *DiaASQ* faces significant challenges due to the fragmented nature of the information and the inherent complex context dependencies in the conversation context.

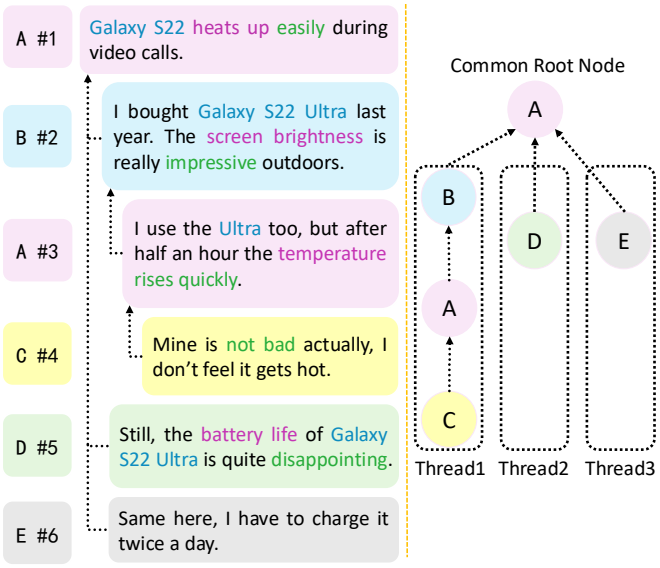

Figure 1: A sample dialogue and its corresponding thread structure.

Type	Target	Aspect	Opinion	Sentiment
Intra-utt	Galaxy S22	heats up	easily	negative
Intra-utt	Galaxy S22 Ultra	screen brightness	impressive	positive
Cross-utt	Galaxy S22 Ultra	temperature	rises quickly	negative
Cross-utt	Galaxy S22 Ultra	temperature	not bad	neutral
Intra-utt	Galaxy S22 Ultra	battery life	disappointing	negative

Table 1: The sentiment quadruple annotations for the dialogue.

*Corresponding author.

To capture the structural details of the conversation, DMIN [Huang *et al.*, 2024] introduced the concept of "discourse thread structure". As shown in Figure 1, the conversation has a highly structured feature, consisting of multiple utterances and their corresponding speakers. These utterances can be decomposed into different semantic threads [Li *et al.*, 2024b; Vedula *et al.*, 2023]. Under this framework, except for the root node, each utterance is closely related to a specific response target, forming a tree-like topological dependency relationship. This complex interaction pattern implies that the flow of sentiment is not only limited by the sequential arrangement of words but also by the topological structure of the conversation. Although the introduction of the thread structure has brought about performance improvements, the existing methods still have difficulty fully leveraging these complex dependencies. Specifically, current models [Li *et al.*, 2024a; Tong *et al.*, 2025; Huang *et al.*, 2024] typically use a general Graph Neural Networks (GCN) to handle the conversation structure, treating the reply-to relations as a simple edge [Schlichtkrull *et al.*, 2018; Veličković *et al.*, 2018]. However, current paradigms usually have two limitations. Firstly, they ignore the semantic isolation between independent threads, inevitably introducing structural noise from irrelevant threads. Secondly, they treat the dynamic conversation as a static graph structure, ignoring the natural temporal order and different speaker identities in the utterances. This failure to implement sequential and speaker-sensitive constraints results in the complex interaction between local context and overall discourse logic not being fully explored [Li *et al.*, 2025].

To capture the relative distances between sentiment elements, Rotary Position Embedding (RoPE) [Su *et al.*, 2024] has been widely adopted in recent *DiaASQ* frameworks [Li *et al.*, 2023; Huang *et al.*, 2024; Li *et al.*, 2024a]. However, existing implementations typically employ a fragmented and cumulative strategy, often restricting entity extraction to the local token context, or simply adding separate attention scores from the token and utterance levels. This token-based modeling introduces a key issue, which we call *Distance Dilution*: in multi-round conversations, verbose utterances expand the distance between logically adjacent turns (e.g., a Q&A pair separated by 50+ tokens). Under high-frequency RoPE rotations, this expanded distance causes the positional correlation to decay prematurely, cutting off semantic connections. Therefore, these mechanisms are difficult to balance both the high sensitivity to local syntax and the long-term retention ability for the global discourse simultaneously.

To address these challenges, we propose the TCDA framework, which integrates explicit topological structure and implicit positioning. Firstly, we introduce the Thread Constraint Directed Acyclic Graph (TC-DAG) to construct an accurate dialogue structure model. Unlike general GCNs that indiscriminately propagate information, TC-DAG sets strict thread-level boundaries. This design effectively suppresses structural noise from irrelevant branches while retaining the logical evolution from the root node to the leaf nodes. Secondly, we propose Discourse-Aware Rotary Position Embedding (D-RoPE) to alleviate the Distance Dilution and overcome the limitations of additive modeling. Unlike the standard encoding method that loosely couples local and global features through linear superposition, D-RoPE constructs a joint semantic-structural embedding. It projects tokens and utterances to independent subspaces and applies topology-adaptive coordinate transformation. This mechanism ensures that fine lexical cues and coarse discourse logic can be deeply integrated before the interaction, enabling accurate interpretation of cross-turn dependencies, regardless of intervening verbosity.

Our contributions can be summarized as follows:

- We propose the Thread Constraint Directed Acyclic Graph (TC-DAG), which employs intra-thread constraints and a fixed root mechanism to suppress structural noise while preserving logical coherence.
- We introduce Discourse-Aware Rotary Position Embedding (D-RoPE), featuring dual-stream projections that decouple micro- and macro-semantics to mitigate distance dilution and align multi-scale distances.
- TCDA achieves SOTA performance. Our code is available at https://github.com/LiXinran6/TCDA.

2 Related Work

2.1 Aspect-Based Sentiment Analysis

Early studies on ABSA mainly focused on simple, isolated sentences with a single structure. Initially, they concentrated on single-element tasks such as aspect extraction [Li *et al.*, 2018] and polarity classification [Li *et al.*, 2021]. To obtain more comprehensive sentiment information, subsequent research shifted to compound tasks, including Aspect-Opinion Pair (AOPE) [Wu *et al.*, 2021] and Triplet Extraction (ASTE) [Chen *et al.*, 2022a; Zhao *et al.*, 2024], which aim to jointly identify aspect terms, opinion terms, and their corresponding polarities. Recently, to provide a comprehensive sentiment picture, the research focus has shifted to Aspect Sentiment Quadruple Prediction (ASQP) [Zhang *et al.*, 2021]. This task extracts the complete (a, c, o, s) quadruple using predefined aspect categories c.

2.2 Conversational Aspect-Based Sentiment Quadruple Analysis

Although the traditional ABSA benchmarks mainly focus on sentence-level [Pontiki *et al.*, 2014; Pontiki *et al.*, 2016], they limit the applicability of existing methods in multi-turn conversation scenarios [Zhang *et al.*, 2023]. To bridge this gap, the *DiaASQ* task was introduced [Li *et al.*, 2023], which employs three parallel attention matrices to explicitly capture the complex inter-utterance correlations. Subsequently, numerous studies further explored this task from different structural perspectives.

H2DT [Li *et al.*, 2024a] employs a heterogeneous attention network and a ternary scorer to enhance the cohesion of quadruples, while DMCA [Li *et al.*, 2024b] and ICMSR [Zhang *et al.*, 2025b] both utilize a multi-scale mechanism - specifically, windows and the SMM module - to capture long-range dependencies and structural features. Specifically, DMIN [Huang *et al.*, 2024] is the first to use GCN and multi-granularity integration to incorporate thread structure,

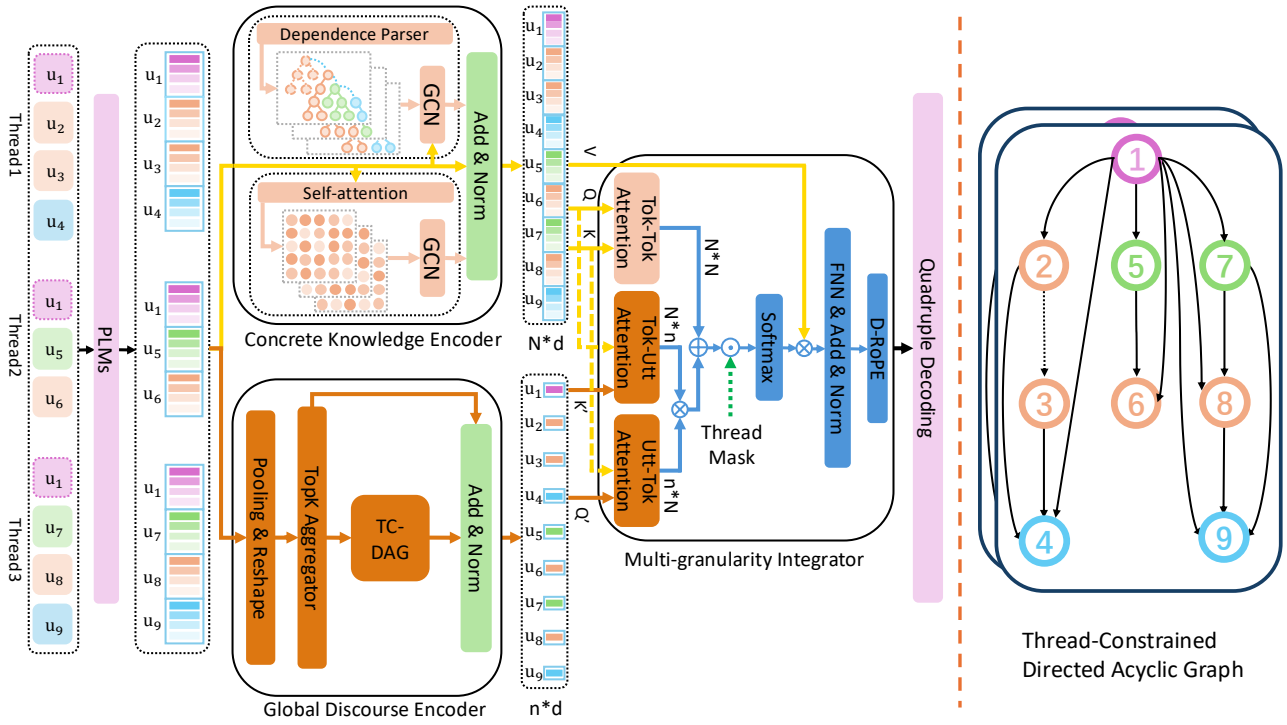

Figure 2: The overall architecture of our proposed TCDA.

enabling token interactions to match the utterance-level discourse. Although CA-DAGNet [Zhang *et al.*, 2025a] constructs a Directed Acyclic Graph [Thost and Chen, 2021; Shen *et al.*, 2021] to capture cross-utterance dependencies, it ignores the inherent thread-based topological constraints. Additionally, recent frameworks [Li *et al.*, 2023; Li *et al.*, 2024a] have integrated RoPE [Su *et al.*, 2024] to encode relative distances within the conversational tree. However, these RoPE implementations are typically limited to encoding the local token context or adopting a fragmented strategy of simple linear superposition, which ignores the differences in frequency scales and cannot alleviate the Distance Dilution caused by verbose utterances.

3 Methodology

We propose TCDA, which combines TC-DAG and D-RoPE. Its overall architecture is shown in Figure 2.

3.1 Problem Definition

In the *DiaASQ* task, each conversation is represented as $D = \{u_1, u_2, \ldots, u_n\}$, along with the reply index set $R = \{l_1, l_2, \ldots, l_n\}$ and the speaker sequence $S = \{s_1, s_2, \ldots, s_n\}$. Here, l_i indicates that the utterance u_i is a direct response to u_{l_i}. Each utterance $u_i = \{w_1, \ldots, w_{m_i}\}$ consists of m_i tokens.

Following the grid tagging framework [Li *et al.*, 2023; Huang *et al.*, 2024], we rephrase the extraction of the quadruple as a unified relation tagging problem. For any pair of words (w_a, w_b) in the flattened dialogue, the model is trained to identify three types of semantic connections:

- **Entity Boundaries** ($y_{ent} \in \{\text{TGT, ASP, OPI}\}$): These labels define the corresponding range by connecting the start and end tokens of the target, aspect, and opinion. For example, the TGT association from "iPhone" to "14" will identify "iPhone 14" as a target entity.
- **Entity Alignment** ($y_{pair} \in \{\text{H2H, T2T}\}$): These relationships link different entities together. Specifically, *head-to-head* (H2H) and *tail-to-tail* (T2T) tags are used to pair the entities, for example, associating the target "iPhone 14" with its corresponding aspect "battery life".
- **Sentiment Polarity** ($y_{pol} \in \{\text{POS, NEG, NEU}\}$): This value indicates the sentiment tendency (positive, negative, or neutral) between the related entities.

For each sub-task, if there is no specific relationship between these tokens, a special label *other* will be assigned to it.

3.2 Textual Feature Extraction

Inspired by DMIN [Huang *et al.*, 2024], each conversation is divided into multiple threads $T_k = \{u'_1, u'_i, u'_{i+1}, \ldots, u'_j\}$, starting from a common root node u'_1, to balance the context window limit on PLM and the discourse interaction. As shown in Figure 1, threads are arranged in sequence and only cross at the root node. Each utterance is formatted as $u'_i = \{[\text{CLS}], u_i, s_i\}$ to incorporate speaker information. The encoding form at the thread level is:

$$H_{T_k} = \{H_1^{u'}, H_i^{u'}, \ldots, H_j^{u'}\} = \text{PLM}(T_k), \qquad (1)$$

where $H_i^{u'} = \{h_i^{\text{cls}}, H_i^u, h_i^s\}$ contains token features $H_i^u \in \mathbb{R}^{m_i \times d}$.

3.3 Dual-scale Contextual Encoding

To simultaneously capture fine-grained semantic cues and coarse-grained discourse structure, we propose a dual-scale encoding framework. This module refines the text representation by performing knowledge enhancement at the thread level and discourse modeling at the conversation level.

Token-level Knowledge Encoding. In order to strike a balance between global and local interactions within the PLM context window, we first perform knowledge enhancement within each individual thread T_k. Following [Huang *et al.*, 2024], we employ a structure called Concrete Knowledge Encoder (CKEncoder), which consists of parallel Syntactic and Semantic GCNs [Kipf and Welling, 2016; Chen *et al.*, 2022b; Zhang *et al.*, 2022; Vaswani *et al.*, 2017]. Specifically, we extract local knowledge features $\tilde{H}_{T_k}$ based solely on the thread-specific context to filter out cross-thread noise:

$$\tilde{H}_{T_k} = \sum_{g \in \{syn, sem\}} \text{GCN}_g(A_g, H_{T_k}), \tag{2}$$

where A_{syn} and A_{sem} respectively represent the thread-level syntactic and semantic adjacency matrices. Subsequently, we aggregate the original features H_{T_k} and the knowledge features $\tilde{H}_{T_k}$ from all threads to reconstruct their global corresponding features H_{tok} and $\tilde{H}_{tok}$ (by averaging the shared root node u'_1). The final enhanced token representation H'_{tok} is obtained through global residual connections and layer normalization:

$$H'_{tok} = \text{LN}(H_{tok} + \tilde{H}_{tok}). \tag{3}$$

Utterance-level Discourse Modeling. Meanwhile, we abstract the original global token-level feature H_{tok} into an utterance-level representation $H_{utt} = \{h_1, \ldots, h_n\}$ through a Top-K aggregator [Huang *et al.*, 2024]. These representations can capture the flow of the conversation, but require powerful structural modeling. Unlike the previous methods that used fully connected graphs, we process H_{utt} using a Thread-Constrained DAG (TC-DAG) to strictly follow the temporal order and replying topology of the conversation. For more details, please refer to Section 3.4.

3.4 Thread-Constrained DAG

To strictly adhere to the dialogue structure and filter out irrelevant information, we propose the Thread-Constrained Directed Acyclic Graph (TC-DAG), which is represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$. Here, $\mathcal{V}$ represents the utterances, and there is a directed edge $(u_j \to u_i)$ only when $j < i$. The relation set $\mathcal{R} = \{0, 1\}$ indicates whether the connected nodes were uttered by the same speaker.

Constructing a Graph through Conversation

A *thread* refers to a sequence within a local conversation branch. To filter out structural noise, TC-DAG employs a retrospective strategy to limit the connection range of edges to be within these threads: each node is connected to the previous utterance that covers ω instances from the same speaker, including all intermediate background information. To ensure global connectivity, when reaching the thread boundary within the window, the connection extends to the root node u_1. This process organizes the conversation into a tree-like DAG (see Figure 3 and Algorithm 1).

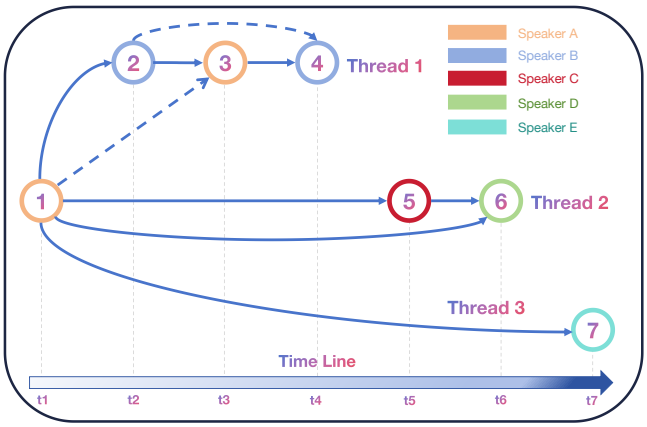

Figure 3: TC-DAG construction ($\omega = 1$). Solid/dashed arrows denote inter- and same-speaker dependencies among chronological utterances. The structure incorporates Global Root Accessibility, allowing nodes in divergent threads (e.g., u_6, u_7) to connect to the global root u_1 under the window constraint.

Structure-Aware Relational Encoding

Based on the constructed TC-DAG and the initial utterance feature H_{utt}, we use a relational GNN to propagate context information along the topological structure. Unlike the standard GNNs, which uniformly aggregates neighbors, our model specifically considers the sequential nature of the conversation and the different dependency types defined in $\mathcal{R}$. Let $\mathbf{h}_i^{(l)}$ represent the hidden state of utterance u_i in the l-th layer, where the input state $\mathbf{h}_i^{(0)}$ corresponds to the vector $\mathbf{h}_i \in H_{utt}$. Since the DAG is strictly arranged in chronological order, we update the nodes from $i = 1$ to n sequentially. This ensures that when calculating u_i, the updated states $\mathbf{h}_j^{(l)}$ of all predecessor utterances $u_j \in \mathcal{N}_i$ (where $j < i$) are already available.

For a specific node u_i, the information aggregation is computed via a relation-aware attention mechanism. The attention coefficient $\alpha_{ij}^{(l)}$ for a neighbor $u_j \in \mathcal{N}_i$ is calculated as:

$$\alpha_{ij}^{(l)} = \text{Softmax}_{j \in \mathcal{N}_i}\left(\mathbf{W}_\alpha^{(l)}\left[\mathbf{h}_j^{(l)} \,\|\, \mathbf{h}_i^{(l-1)}\right]\right) \tag{4}$$

where $\|$ denotes concatenation. The context vector $\mathbf{m}_i^{(l)}$ is then derived by:

$$\mathbf{m}_i^{(l)} = \sum_{j \in \mathcal{N}_i} \alpha_{ij}^{(l)} \mathbf{W}_{r_{ij}}^{(l)} \mathbf{h}_j^{(l)} \tag{5}$$

where $\mathbf{W}_{r_{ij}}^{(l)} \in \{\mathbf{W}_0^{(l)}, \mathbf{W}_1^{(l)}\}$ is a relation-specific projection matrix selected based on whether u_i and u_j share the same speaker ($r_{ij} \in \mathcal{R}$). This allows the model to differentially weigh intra-speaker and inter-speaker dependencies.

In order to effectively integrate the aggregated contextual information with the node's own historical records, we adopt a dual gated update mechanism [Shen *et al.*, 2021]. Specifically, we employ two parallel GRU units to capture complementary information flows. The *node update unit* (GRU_H) uses the context as guidance to update the node's state, while

Algorithm 1 Building a Thread-Constrained DAG

Require: Dialogue $\{u_1, \ldots, u_N\}$, speaker $P(\cdot)$, thread mapping $T(\cdot)$, window size ω.
Ensure: Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$.

$\mathcal{V} \leftarrow \{u_1, \ldots, u_N\}, \mathcal{E} \leftarrow \emptyset, \mathcal{R} \leftarrow \{0, 1\}$
for $i = 2$ **to** N **do**
  $c \leftarrow 0, \quad \tau \leftarrow i - 1$
  $S_i \leftarrow$ Start index of thread $T(u_i)$
  while $\tau \geq S_i$ **and** $c < \omega$ **do**
    $r \leftarrow (P(u_\tau) = P(u_i))?1 : 0$
    $\mathcal{E} \leftarrow \mathcal{E} \cup \{(u_\tau, u_i, r)\}$
    if $r = 1$ **then**
      $c \leftarrow c + 1$
    end if
    $\tau \leftarrow \tau - 1$
  end while
  if $c < \omega$ **and** $S_i > 1$ **then**
    $r \leftarrow (P(u_1) = P(u_i))?1 : 0$
    $\mathcal{E} \leftarrow \mathcal{E} \cup \{(u_1, u_i, r)\}$
  end if
end for
return $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$

the *context update unit* (GRU_C) models the evolution of the context:

$$\tilde{\mathbf{h}}_i^{(l)} = \text{GRU}_H(\mathbf{h}_i^{(l-1)}, \mathbf{m}_i^{(l)}) \tag{6}$$

$$\mathbf{c}_i^{(l)} = \text{GRU}_C(\mathbf{m}_i^{(l)}, \mathbf{h}_i^{(l-1)}) \tag{7}$$

Here, the inputs and hidden states are logically swapped between the two GRUs to maximize feature interaction. Finally, the updated representation for node u_i at layer l is obtained by summing the outputs:

$$\mathbf{h}_i^{(l)} = \tilde{\mathbf{h}}_i^{(l)} + \mathbf{c}_i^{(l)} \tag{8}$$

Finally, we extract the node states $H_{utt}^{(L)} = \{\mathbf{h}_1^{(L)}, \ldots, \mathbf{h}_n^{(L)}\}$ from the last layer L and apply a residual connection followed by layer normalization to yield the final global representations:

$$H'_{utt} = \text{LN}(H_{utt} + H_{utt}^{(L)}). \tag{9}$$

3.5 Global-Local Interaction and Discourse-Aware Position Encoding

After obtaining the global structure-aware representation H'_{utt} through the TC-DAG module, our aim is to reintegrate this global background information into the token-level features and enhance the position sensitivity.

Global-Local Interaction

To bridge the gap between the coarse-grained discourse structure and the fine-grained token features, we employ the cross-attention mechanism. Token representation H'_{tok} is used as the *query*, while the global utterance representation H'_{utt} serves as the *key* and *value*, enabling tokens to focus on the relevant discourse context and generate the comprehensive representation H_{final}.

Algorithm 2 D-RoPE-Enhanced Attention

Require: Queries/Keys $\mathbf{Q}, \mathbf{K}$ for tok/utt; Indices P_{tok}, P_{utt}.
Ensure: Score matrix $\mathbf{S}$.

Projection: $\mathbf{V}^{mic/mac} \leftarrow \text{Linear}_{mic/mac}(\mathbf{V}^{tok/utt})$ for $\mathbf{V} \in \{\mathbf{Q}, \mathbf{K}\}$
Coordinate Transform: For pair (i, j), let $\sigma_{ij} = -1$ if threads diverge else 1. Set $\hat{p}^{(j)} \leftarrow \sigma_{ij} P^{(j)}$.
Dual-Scale Rotation:
  $\tilde{\mathbf{q}}_i \leftarrow \text{Concat}(\mathcal{R}(\mathbf{q}_i^{mic}, P_{tok}^{(i)}), \mathcal{R}(\mathbf{q}_i^{mac}, P_{utt}^{(i)}))$
  $\tilde{\mathbf{k}}_j \leftarrow \text{Concat}(\mathcal{R}(\mathbf{k}_j^{mic}, \hat{p}_{tok}^{(j)}), \mathcal{R}(\mathbf{k}_j^{mac}, \hat{p}_{utt}^{(j)}))$
return $\mathbf{S}$ where $\mathbf{S}_{ij} = \tilde{\mathbf{q}}_i^\top \tilde{\mathbf{k}}_j$

Data	Set	D	U	Q_{tot}	Q_{int}	Q_{cro}
ZH	Train	800	5,947	4,607	3,594	1,013
	Valid	100	748	577	440	137
	Test	100	757	558	433	125
	Total	1,000	7,452	5,742	4,467	1,275
EN	Train	800	5,947	4,414	3,442	972
	Valid	100	748	555	423	132
	Test	100	757	545	422	123
	Total	1,000	7,452	5,514	4,287	1,227

Table 2: Statistics of ZH and EN datasets.

Discourse-Aware Rotary Position Embedding (D-RoPE)
To alleviate the inherent *Distance Dilution* phenomenon in the RoPE strategy, our D-RoPE method explicitly separates the semantic granularity into independent subspaces and fuses them before interaction.

Dual-Scale Semantic-Structural Projection. We decompose the integrated representation H_{final} into parallel tokens ($\mathbf{h}_{tok}$) and utterances ($\mathbf{h}_{utt}$) streams, and then project them onto separate subspaces:

$$\mathbf{q}_i^{mic} = \mathbf{W}_{mic}\mathbf{h}_{tok,i}, \quad \mathbf{q}_i^{mac} = \mathbf{W}_{mac}\mathbf{h}_{utt,i} \tag{10}$$

$$\mathbf{k}_j^{mic} = \mathbf{W}_{mic}\mathbf{h}_{tok,j}, \quad \mathbf{k}_j^{mac} = \mathbf{W}_{mac}\mathbf{h}_{utt,j} \tag{11}$$

where $\mathbf{W}_{mic}$ and $\mathbf{W}_{mac}$ are learnable matrices that separate local syntactic cues from the global discourse semantics.

Topology-Adaptive Rotary Encoding. We employ RoPE method with different base frequencies to encode the topological structure. While maintaining the standard relative position property $\tilde{\mathbf{q}}^\top \tilde{\mathbf{k}} = \mathbf{q}^\top \mathcal{R}(p_q - p_k)\mathbf{k}$, we introduce a Topology-Adaptive Coordinate Transformation that is applicable at both the micro and macro levels:

1. **Micro-RoPE (Token Level)**: With a standard frequency $\theta_{mic} = 10000$, we define the token index p_{tok} as the cumulative topological distance starting from the global root node [Li *et al.*, 2023]. To make the subtraction mechanism of RoPE compatible with the addition distance (i.e., $p_{tok}^{(i)} + p_{tok}^{(j)}$ between different branch threads), we apply the coordinate sign inversion:

$$\hat{p}_{tok}^{(j)} = \begin{cases} p_{tok}^{(j)} & \text{if } x_i, x_j \text{ in same thread} \\ -p_{tok}^{(j)} & \text{if } x_i, x_j \text{ in divergent threads} \end{cases} \tag{12}$$

Data	Model	Pair Extraction (F1)			Quadruple (F1)	
		T-A	T-O	A-O	Micro	Ident.
ZH	MVQPN* [Li *et al.*, 2023]	50.07	50.40	51.91	35.68	41.37
	H2DT* [Li *et al.*, 2024a]	50.00	48.20	52.56	39.85	43.03
	DMCA [Li *et al.*, 2024b]	56.88	51.70	52.80	42.68	45.36
	DMIN* [Huang *et al.*, 2024]	<u>57.61</u>	51.18	**55.58**	<u>43.29</u>	<u>46.02</u>
	CA-DAGNet [Zhang *et al.*, 2025a]	**58.76**	51.45	52.65	42.47	45.44
	IFusionQuad [Jiang *et al.*, 2025]	54.68	51.81	50.04	41.53	44.56
	ICMSR [Zhang *et al.*, 2025b]	56.69	**52.53**	54.59	42.55	45.20
	TCDA (Ours)	57.47	<u>52.24</u>	<u>55.46</u>	**44.35**	**46.23**
EN	MVQPN* [Li *et al.*, 2023]	48.60	48.31	50.05	35.62	38.86
	H2DT* [Li *et al.*, 2024a]	48.41	49.14	51.87	38.76	41.95
	DMCA [Li *et al.*, 2024b]	53.08	50.99	**52.40**	37.96	41.00
	DMIN* [Huang *et al.*, 2024]	53.72	52.40	52.22	38.14	41.85
	CA-DAGNet [Zhang *et al.*, 2025a]	54.04	51.88	50.97	37.87	41.52
	IFusionQuad [Jiang *et al.*, 2025]	52.65	51.82	51.94	35.96	41.49
	ICMSR [Zhang *et al.*, 2025b]	<u>54.23</u>	<u>52.67</u>	51.61	<u>39.36</u>	**44.06**
	TCDA (Ours)	**55.40**	**52.90**	<u>52.29</u>	**39.69**	<u>42.12</u>

Table 3: Performance on DiaASQ. T/A/O denote Target, Aspect, and Opinion; Ident. is Identification F1. Results with * are our reproductions; others are cited. Best and second-best results are bolded and underlined.

This transformation enables $\mathcal{R}(p^{(i)}_{tok} - \hat{p}^{(j)}_{tok})$ to accurately encode the topological path lengths between different threads, while preserving the linear relative distances within the same thread.

2. **Macro RoPE (Utterance Level)**: Relying solely on token indexing can lead to *distance dilution*, where verbose utterances increase the distance and disrupt the semantic connections under high-frequency rotation. To alleviate this, we introduce Macro-RoPE, using utterance-level index p_{utt}, with the base frequency $\theta_{mac} = 100$ reduced. This transformation preserves strong attention on logical dependencies:

$$\hat{p}^{(j)}_{utt} = \begin{cases} p^{(j)}_{utt} & \text{if } x_i, x_j \text{ in same thread} \\ -p^{(j)}_{utt} & \text{if } x_i, x_j \text{ in divergent threads} \end{cases} \tag{13}$$

This ensures constant turn-level distances, serving as a robust discourse anchor.

Fusion. We construct a unified feature vector by concatenating the rotation embeddings of the two subspaces:

$$\tilde{\mathbf{q}}_i = [\tilde{\mathbf{q}}^{mic}_i \,\|\, \tilde{\mathbf{q}}^{mac}_i] \tag{14}$$

$$\tilde{\mathbf{k}}_j = [\tilde{\mathbf{k}}^{mic}_j \,\|\, \tilde{\mathbf{k}}^{mac}_j] \tag{15}$$

Here, $[\cdot \,\|\, \cdot]$ represents concatenation. Then, the topological adaptive score is calculated through the dot product:

$$\text{Score}(x_i, x_j) = \tilde{\mathbf{q}}^{\top}_i \tilde{\mathbf{k}}_j \tag{16}$$

This ensures dual-scale semantic and positional consistency.

3.6 Quadruple Decoding and Learning

To isolate the semantic influence, we project the H_{final} item into three task-specific spaces (S_{ent}, S_{rel}, S_{pol}). We apply D-RoPE to each grid g to derive topology-adaptive probabilities by Softmax:

$$P(y^g_{ij}|x_i, x_j) = \text{Softmax}(\text{Score}_g(x_i, x_j)) \tag{17}$$

We minimize weighted cross-entropy loss:

$$\mathcal{L} = -\sum_g \sum_{i,j} \alpha^g_{ij} \log P(y^g_{ij}|x_i, x_j) \tag{18}$$

where y^g_{ij} represents the true label, while α^g_{ij} denotes the category weight.

4 Experiments and Analysis

4.1 Dataset and Implementation Details

Dataset. We conduct experiments on the Chinese (ZH) and English (EN) datasets [Li *et al.*, 2023]. The detailed statistics are presented in Table 2.

Implementation Details. Following existing methods, we use RoBERTa-Large [Liu *et al.*, 2019] and Chinese-RoBERTa-wwm-ext-base [Cui *et al.*, 2019] as backbones for EN and ZH, with Top-K ratios λ of 0.5 and 0.8, respectively. Both syntactic and semantic GCNs consist of 3 layers, while the TC-DAG has 2 layers. We employ a sliding window of size $w = 3$. We train with a batch size of 2 and a 0.1 dropout rate. The AdamW optimizer is used with learning rates of 1e-5 for PLMs and 1e-4 for other parameters. All experiments are conducted on a single NVIDIA GeForce RTX 4090 GPU. All results, including baseline comparisons and ablation studies, are reported as the average of five independent runs to ensure statistical significance.

4.2 Baselines

We compare TCDA against several state-of-the-art baselines: MVQPN [Li *et al.*, 2023] (the pioneering grid-tagging baseline), H2DT [Li *et al.*, 2024a], DMCA [Li *et al.*, 2024b],

DMIN [Huang *et al.*, 2024], CA-DAGNet [Zhang *et al.*, 2025a], IFusionQuad [Jiang *et al.*, 2025] and ICMSR [Zhang *et al.*, 2025b].

4.3 Main Results

Table 3 shows that TCDA achieves SOTA or competitive performance across all benchmarks.

4.4 Ablation Study

To assess the contribution of each component, we compare TCDA with three variants: (1) **w/o TC-DAG**, replacing the thread-constrained topology with the standard reply-based GCN; (2) **w/o D-RoPE**, replacing the Discourse-Aware positioning with the standard RoPE; (3) **w/o Both**, removing both modules.

Table 4 shows that removing any component degrades performance, with the sharpest decline when both are absent. This confirms that TC-DAG and D-RoPE provide complementary benefits in filtering noise and addressing distance dilution.

Variant	ZH		EN	
	F1	Δ	F1	Δ
TCDA (Full)	**44.35**	-	**39.69**	-
w/o TC-DAG	43.78	-0.57	38.78	-0.91
w/o D-RoPE	43.74	-0.61	38.65	-1.04
w/o Both	43.29	-1.06	38.14	-1.55

Table 4: Ablation results (Micro F1).

4.5 Further Analysis

Parameter Sensitivity. We investigate the impact of the TC-DAG layer L and the speaker window size w on the performance, as shown in Table 5. All other hyperparameters (including the standard RoPE baseline values) are kept constant to ensure the fairness of the comparison.

Layers (L)	F1	Window (w)	F1
$L = 1$	43.41	$w = 1$	42.34
$L = 2$	43.74	$w = 2$	43.74
$L = 3$	43.18	$w = 3$	43.74
$L = 4$	43.26	$w = 4$	43.74

Table 5: Parameter sensitivity on ZH dataset.

The best performance can be achieved when $L = 2$, and increasing L can lead to a decrease in performance due to the over-smoothing effect. Dense connection ($w \geq 2$) is always superior to sparse connection ($w = 1$), as it can facilitate direct information transmission from the root sentence to subsequent nodes, thereby maintaining the overall discourse intention in the case of long-distance attenuation. Performance saturates at $w \geq 2$ as the window size often exceeds the actual thread length.

Generality of D-RoPE. To verify the universality of D-RoPE, we replace the standard RoPE with our D-RoPE in the competitive benchmark models (i.e., MVQPN and DMIN). As shown in Table 6, D-RoPE consistently improves performance across different architectures and languages. Notably, for MVQPN, its micro F1 value increases by 1.84% on the ZH dataset and 0.80% on the EN dataset. This significant improvement indicates that D-RoPE effectively overcomes the limitations of the base model in capturing multi-scale positional dependencies. Moreover, the continuous improvement of DMIN further confirms that D-RoPE is a robust, model-independent plugin that can alleviate the Distance Dilution problem.

Base Model	ZH		EN	
	F1	Gain	F1	Gain
MVQPN [Li *et al.*, 2023]	35.68	-	35.62	-
+ D-RoPE	37.52	↑ 1.84	36.42	↑ 0.80
DMIN [Huang *et al.*, 2024]	43.29	-	38.14	-
+ D-RoPE	43.78	↑ 0.49	38.78	↑ 0.64

Table 6: Generality of D-RoPE on different baselines (Micro F1).

Effectiveness of TC-DAG Structure. As shown in Table 7, to verify the necessity of our topological consistency design, we compare the proposed TC-DAG with two structural variants: (1) Reply-GCN, which only builds an undirected graph based on reply dependencies, ignoring speaker relationships and directionality; (2) Standard DAG, which follows the chronological order and distinguishes edges by speakers. We use the standard RoPE method in all variants.

Without thread isolation (standard DAG), unrelated thread interference degrades performance below the reply-GCN baseline on the EN dataset. Conversely, TC-DAG eliminates this interference by integrating chronological order with strict topology, achieving superior results across all metrics.

Graph Structure	ZH	EN
Reply-GCN (Undirected)	43.29	38.14
Standard DAG	43.48	37.57
TC-DAG (Ours)	**43.74**	**38.65**

Table 7: Impact of graph construction strategies (Micro F1).

5 Conclusion

We propose the TCDA framework to address the structural noise and scale mismatch issues in DiaASQ. We introduce TC-DAG to filter out irrelevant branches by implementing topological constraints, and introduce D-RoPE to solve the Distance Dilution by aligning the semantic granularity with the hierarchical structure of the separated subspace. TCDA achieves SOTA results in two benchmarks. The generalization ability of D-RoPE further highlights its potential as a model-agnostic plugin suitable for a wider range of multi-turn dialogue tasks.

Acknowledgments

This work was supported by the National Natural Science Foundation of China Project (No. 62372078).

References

[Chen *et al.*, 2022a] Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2974–2985, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[Chen *et al.*, 2022b] Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2974–2985, Dublin, Ireland, May 2022. Association for Computational Linguistics.

[Cui *et al.*, 2019] Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514, 2019.

[Huang *et al.*, 2024] Peijie Huang, Xisheng Xiao, Yuhong Xu, and Jiawei Chen. DMIN: A discourse-specific multi-granularity integration network for conversational aspect-based sentiment quadruple analysis. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Findings of the Association for Computational Linguistics: ACL 2024*, pages 16326–16338, Bangkok, Thailand, August 2024. Association for Computational Linguistics.

[Jiang *et al.*, 2025] Haoyu Jiang, Xiaoliang Chen, Duoqian Miao, Hongyun Zhang, Xiaolin Qin, Xu Gu, and Peng Lu. 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, 2025.

[Kipf and Welling, 2016] Thomas Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *ArXiv*, abs/1609.02907, 2016.

[Li *et al.*, 2018] Xin Li, Lidong Bing, Piji Li, Wai Lam, and Zhimou Yang. Aspect term extraction with history attention and selective transformation. In *International Joint Conference on Artificial Intelligence*, 2018.

[Li *et al.*, 2021] Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. Dual graph convolutional networks for aspect-based sentiment analysis. 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 (Volume 1: Long Papers)*, pages 6319–6329, Online, August 2021. Association for Computational Linguistics.

[Li *et al.*, 2023] Bobo Li, Hao Fei, Fei Li, Yuhan Wu, Jinsong Zhang, Shengqiong Wu, Jingye Li, Yijiang Liu, Lizi Liao, Tat-Seng Chua, and Donghong Ji. DiaASQ: A benchmark of conversational aspect-based sentiment quadruple analysis. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 13449–13467, Toronto, Canada, July 2023. Association for Computational Linguistics.

[Li *et al.*, 2024a] Bobo Li, Hao Fei, Lizi Liao, Yu Zhao, Fangfang Su, Fei Li, and Donghong Ji. Harnessing holistic discourse features and triadic interaction for sentiment quadruple extraction in dialogues. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'24/IAAI'24/EAAI'24. AAAI Press, 2024.

[Li *et al.*, 2024b] Yuqing Li, Wenyuan Zhang, Binbin Li, Siyu Jia, Zisen Qi, and Xingbang Tan. Dynamic multi-scale context aggregation for conversational aspect-based sentiment quadruple analysis. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 11241–11245, 2024.

[Li *et al.*, 2025] Xinran Li, Xiujuan Xu, and Jiaqi Qiao. Long-short distance graph neural networks and improved curriculum learning for emotion recognition in conversation. In *Proceedings of the 28th European Conference on Artificial Intelligence (ECAI 2025)*, volume 413 of *Frontiers in Artificial Intelligence and Applications*, pages 4033–4040. IOS Press, 2025.

[Liu *et al.*, 2019] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *ArXiv*, abs/1907.11692, 2019.

[Mao *et al.*, 2022] Yue Mao, Yi Shen, Jingchao Yang, Xiaoying Zhu, and Longjun Cai. Seq2Path: Generating sentiment tuples as paths of a tree. In Smaranda Muresan, Preslav Nakov, and Aline Villavicencio, editors, *Findings of the Association for Computational Linguistics: ACL 2022*, pages 2215–2225, Dublin, Ireland, May 2022. Association for Computational Linguistics.

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

[Pontiki *et al.*, 2016] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub,

Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. SemEval-2016 task 5: Aspect based sentiment analysis. In Steven Bethard, Marine Carpuat, Daniel Cer, David Jurgens, Preslav Nakov, and Torsten Zesch, editors, *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June 2016. Association for Computational Linguistics.

[Schlichtkrull *et al.*, 2018] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehwish Alam, editors, *The Semantic Web*, pages 593–607, Cham, 2018. Springer International Publishing.

[Shen *et al.*, 2021] Weizhou Shen, Siyue Wu, Yunyi Yang, and Xiaojun Quan. Directed acyclic graph network for conversational emotion recognition. 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 (Volume 1: Long Papers)*, pages 1551–1560, Online, August 2021. Association for Computational Linguistics.

[Su *et al.*, 2024] Jianlin Su, Murtadha Ahmed, Yu Lu, Shengfeng Pan, Wen Bo, and Yunfeng Liu. Roformer: Enhanced transformer with rotary position embedding. *Neurocomputing*, 568:127063, 2024.

[Thost and Chen, 2021] Veronika Thost and Jie Chen. Directed acyclic graph neural networks. In *International Conference on Learning Representations*, 2021.

[Tong *et al.*, 2025] Zeliang Tong, Wei Wei, Xiaoye Qu, Rikui Huang, Zhixin Chen, and Xingyu Yan. Multi-level association refinement network for dialogue aspect-based sentiment quadruple analysis. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar, editors, *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 14035–14057, Vienna, Austria, July 2025. Association for Computational Linguistics.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Neural Information Processing Systems*, 2017.

[Vedula *et al.*, 2023] Nikhita Vedula, Marcus Collins, and Oleg Rokhlenko. Disentangling user conversations with voice assistants for online shopping. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '23, page 1939–1943, New York, NY, USA, 2023. Association for Computing Machinery.

[Veličković *et al.*, 2018] Petar Veličković, Adriana Casanova, Pietro Liò, Guillem Cucurull, Adriana Romero, and Yoshua Bengio. Graph attention networks. 2018.

[Wu *et al.*, 2021] Shengqiong Wu, Hao Fei, Yafeng Ren, Donghong Ji, and Jingye Li. Learn from syntax: Improving pair-wise aspect and opinion terms extractionwith rich syntactic knowledge. In *International Joint Conference on Artificial Intelligence*, 2021.

[Zhang *et al.*, 2021] Wenxuan Zhang, Yang Deng, Xin Li, Yifei Yuan, Lidong Bing, and Wai Lam. Aspect sentiment quad prediction as paraphrase generation. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9209–9219, Online and Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.

[Zhang *et al.*, 2022] Zheng Zhang, Zili Zhou, and Yanna Wang. SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4916–4925, Seattle, United States, July 2022. Association for Computational Linguistics.

[Zhang *et al.*, 2023] Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, 35(11):11019–11038, 2023.

[Zhang *et al.*, 2025a] Qiang Zhang, Jie Zeng, Runze Zhang, and Dong Cui. Context-aware directed acyclic graph network for conversational aspect-based sentiment quadruple analysis. *IEEE Access*, 13:154823–154832, 2025.

[Zhang *et al.*, 2025b] Yu Zhang, Zhaoman Zhong, and Huihui Lv. Inter-sentence context modeling and structure-aware representation enhancement for conversational sentiment quadruple extraction. In Christos Christodoulopoulos, Tanmoy Chakraborty, Carolyn Rose, and Violet Peng, editors, *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 17149–17159, Suzhou, China, November 2025. Association for Computational Linguistics.

[Zhao *et al.*, 2024] Xiaowei Zhao, Yong Zhou, and Xiujuan Xu. Dual encoder: Exploiting the potential of syntactic and semantic for aspect sentiment triplet extraction. In Nicoletta Calzolari, Min-Yen Kan, Veronique Hoste, Alessandro Lenci, Sakriani Sakti, and Nianwen Xue, editors, *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 5401–5413, Torino, Italia, May 2024. ELRA and ICCL.