The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

Harnessing Holistic Discourse Features and Triadic Interaction for Sentiment Quadruple Extraction in Dialogues

Bobo Li[1], Hao Fei[2], Lizi Liao[3], Yu Zhao[4], Fangfang Su[1], Fei Li[1], Donghong Ji[1]*

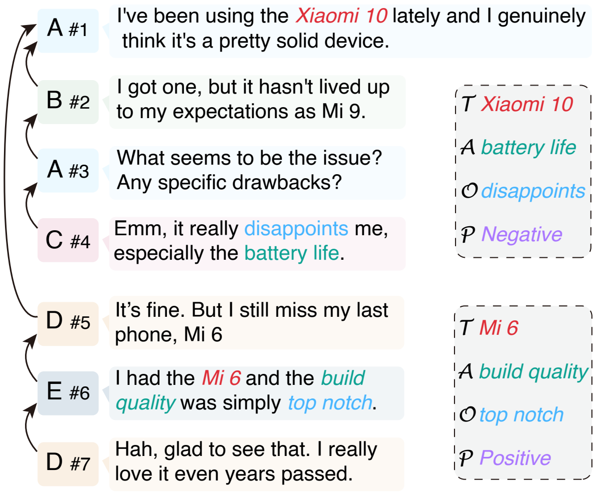
[1] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of
Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China
[2] School of Computing, National University of Singapore, Singapore
[3] School of Computing and Information Systems, Singapore Management University, Singapore
[4] College of Intelligence and Computing, Tianjin University, Tianjin, China
{boboli, fangsu, lifei_csnlp, dhji}@whu.edu.cn, haofei37@nus.edu.sg, lzliao@smu.edu.sg, zhaoyucs@tju.edu.cn

Abstract

Dialogue Aspect-based Sentiment Quadruple (DiaASQ) is a newly-emergent task aiming to extract the sentiment quadruple (i.e., targets, aspects, opinions, and sentiments) from conversations. While showing promising performance, the prior DiaASQ approach unfortunately falls prey to the key crux of DiaASQ, including insufficient modeling of discourse features, and lacking quadruple extraction, which hinders further task improvement. To this end, we introduce a novel framework that not only capitalizes on comprehensive discourse feature modeling, but also captures the intrinsic interaction for optimal quadruple extraction. On the one hand, drawing upon multiple discourse features, our approach constructs a token-level heterogeneous graph and enhances token interactions through a heterogeneous attention network. We further propose a novel triadic scorer, strengthening weak token relations within a quadruple, thereby enhancing the cohesion of the quadruple extraction. Experimental results on the DiaASQ benchmark showcase that our model significantly outperforms existing baselines across both English and Chinese datasets. Our code is available at https://bit.ly/3v27pqA.

Introduction

Aspect-based Sentiment Analysis (ABSA) has garnered considerable research attention in the field of affective computing (Zhang et al. 2023b). The core of ABSA involves extracting opinions or sentiment preferences towards specific aspects from text, forming tasks such as target-oriented opinion words extraction (Fan et al. 2019), aspect sentiment triplet extraction (Peng et al. 2020), and aspect sentiment quad prediction (Zhang et al. 2021a). Regardless of the diverse forms of current ABSA research, the primary focus remains limited to individual text pieces, such as online reviews. However, in daily interactions, e.g., on social media, people tend to convey their fine-grained sentiments in the dialogue format. Due to the intricate nature of dialogue discourse, traditional ABSA techniques developed for single-text might fail to fully capture the richness and depth of sentiments within them (Cai, Xia, and Yu 2021; Zhang et al. 2021a). This potential has sparked growing research interest in ABSA for dialogues, enabling real-world applications.



I've been using the *Xiaomi 10* lately and I genuinely think it's a pretty solid device.

I got one, but it hasn't lived up to my expectations as Mi 9.

What seems to be the issue? Any specific drawbacks?

Emm, it really disappoints me, especially the battery life.

It's fine. But I still miss my last phone, Mi 6

I had the *Mi 6* and the *build quality* was simply *top notch*.

Hah, glad to see that. I really love it even years passed.

$\mathcal{T}$ *Xiaomi 10*
$\mathcal{A}$ *battery life*
$\mathcal{O}$ *disappoints*
$\mathcal{P}$ *Negative*
$\mathcal{T}$ *Mi 6*
$\mathcal{A}$ *build quality*
$\mathcal{O}$ *top notch*
$\mathcal{P}$ *Positive*

Figure 1: An example of aspect-based quadruple extraction in conversation. The item '$\mathcal{T}$', '$\mathcal{A}$', '$\mathcal{O}$', and '$\mathcal{P}$' denote *target*, *aspect*, *opinion*, and *polarity*, respectively. The arraw→ represents the reply relationship in conversation.

In light of the promising applications, Li et al. (2023a) pioneered ABSA analysis in dialogue, introducing the task of DiaASQ. As illustrated in Figure 1, given a multi-party, multi-turn dialogue, the task aims to extract the quadruple formed by *target, aspect, opinion*, and *sentiment polarity*. Meanwhile, to benchmark the DiaASQ task, Li et al. (2023a) proposed a word-pair-based framework by following prior quadruple extraction (Zhang et al. 2021a). Although exhibiting promising performance, their model still falls short of two key limitations, in terms of two crux of DiaASQ, i.e., discourse modeling and triadic quadruple interaction. First, compared to previous ABSA in the individual texts, sentiment-related content of DiaASQ is conveyed through dialogue flow, necessitating an efficient understanding of dialogue discourse to extract quadruples. Besides, the extraction process of DiaASQ, which involves detecting three items and subsequently assigning a sentiment label to them, presents an urgent demand for triadic interaction between

*Corresponding author

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

quadruples within complex conversational contexts.

Insufficient Discourse Modeling Discourse features play a pivotal role in the DiaASQ task, especially in cross-utterance extraction, as exemplified by the first quadruple in Figure 1. In real-life conversations, individuals frequently shift focus and express opinions with fluidity. As a result, it becomes imperative to analyze discourse features and to thoroughly understand dialogue-level structures. Regrettably, Li et al. (2023a) fail to adequately leverage these structures, i.e., they employed three separate attention matrices to capture the relationships between each utterance pair on the same thread, the same speaker, and the reply relation. Yet, such a detached manner neglects the comprehensive nature of discourse features, resulting in a potentially myopic understanding. Considering the scenario illustrated in Figure 1, a sole focus on the same-thread interaction could mistakenly extract the false quadruple ('Mi 9', 'battery life', 'disappointed', *negative*) due to the proximity of the 2nd and 4th utterances. Instead, it is much preferable to integrate both the reply and speaker relationships effectively, treating them as pivotal discourse features to enhance quadruple extraction.

Neglecting Quadruple Cohesion Much of the prior research has sought to streamline quadruple extraction by dividing it into three distinct pair extractions (Zhang et al. 2021a; Li et al. 2023a). Specifically, when all three pairs involving the target ($\mathcal{T}$), aspect ($\mathcal{A}$), and opinion ($\mathcal{O}$) are deemed valid, they collectively, along with a sentiment polarity, constitute a quadruple. While this method may seem to simplify the task, it largely weakens the cohesion of the quadruple, making the overall extraction quality dependent on the performance of the weakest pair. As can be exemplified, in a study presented in Li et al. (2023a), some models performed well on certain pairs but had limited overall success due to poor results with other pairs. A potential solution is to consider additional tokens in higher order during pair extraction, especially those within the quadruple, to provide more comprehensive viewpoints of the quadruple. Considering an example depicted in Figure 1, a direct correlation solely between 'Mi 6' and 'top notch' tends to overlook nuanced facets of 'Mi 6', thus rendering the associated evaluation (i.e., 'top notch') somewhat ambiguous. By integrating the intermediary aspect 'build quality', the evaluation becomes precisely anchored, indicating that 'top notch' specifically appraises the 'build quality' of 'Mi 6'.

To address the aforementioned challenges, we introduce a unified model named Harness Holist Discourse Features and Triadic Interaction (H2DT) for effective quadruple extraction. In the feature interaction stage: We introduce an R-S (Reply & Speaker) heterogeneous graph, aiming to holistically capture the intricate relationships between tokens in dialogues. Through the construction of a meta-path between tokens, both reply and speaker role information is integrated, enriching the discourse feature within the token representation. In the label encoding stage, We transform the quadruple into token pair labels, producing two matrices: the entity matrix and the relation matrix. Specifically, within the relation matrix, we employ a triadic interaction mechanism. This mechanism enhances the token pair interaction by integrating a third token's representation, offering a comprehensive perspective for quadruple extraction. In the inference stage, we predict the labels between token pairs, and the quadruple can readily be derived from the two matrices.

We conducted extensive experiments on the DiaASQ dataset to evaluate our H2DT model. Notably, our H2DT model outperformed baselines on both the Chinese and English datasets, achieving improvements of 5.4 and 5.7 in F1 scores, respectively. In-depth analyses highlight the efficacy of our heterogeneous graph for cross-utterance quadruple extraction. Additionally, the triadic interaction that considers a third token during token-pair label prediction, enhances the overall quadruple quality. Moreover, our analysis shows that the H2DT integrates the triadic interaction across the sequence without incurring substantial computational complexity, underscoring its efficiency and streamlined design.

Overall, this paper revisits the key bottlenecks of DiaASQ and makes contributions in the following aspects:

- We propose a holistic discourse feature modeling mechanism, underpinned by a heterogeneous attention network, to refine token representation for the DiaASQ task.
- We introduce an efficient triadic scorer to incorporate a third token into the token-pair interaction process, thereby enhancing the coherence of the quadruple.
- Experimental results demonstrate the superior performance of our proposed model, pushing the current arts of the DiaASQ task.

Related Work

Aspect-based Sentiment Analysis Aspect-based sentiment analysis (ABSA) has become a pivotal research area within the field of affective computing (Picard 2000; Zhang et al. 2023b; Fei et al. 2020), with a significant foundation laid using the SemEval dataset (Pontiki et al. 2014, 2015, 2016). Broadly, ABSA methodologies fall into two primary categories (Zhang et al. 2023b; Fei et al. 2022a). The first, **single task**, emphasizes individual components like aspect terms, opinion terms, categories, and sentiment polarities for an aspect. Approaches span a range from token-level classification (Liu, Joty, and Meng 2015; Wang et al. 2016) to machine reading comprehension (Mao et al. 2021; Gao et al. 2021) and seq-to-seq models (Yang et al. 2020; Yan et al. 2021). Contrastingly, the **compound task** delves into multiple-item extraction, such as aspect-opinion pairing or aspect sentiment quadruple extraction. Innovations in this field have produced techniques like span-based models (Wu et al. 2021; Xu, Chia, and Bing 2021; Zhang et al. 2023a), machine reading comprehension-based models (Chen et al. 2021; Mao et al. 2021; Gao et al. 2021), and generative-based models (Zhang et al. 2021b,a; Peper and Wang 2022). Highlighting recent progress, Li et al. (2023a) showcased ABSA at the dialogue level, underscoring challenges in dialogue comprehension and quadruple extraction. Thus, this paper's focus shifts to extraction within DiaASQ.

Discoure Feature Modeling Dialogues, with their inherent complexities such as utterances, speaker roles, reply relationships, and threads, pose a greater understanding challenge compared to flat document texts. Numerous studies have delved into modeling these dialogue-specific features. Studies such as (Wei, Xu, and Mao 2019) have probed into the sequential nature of dialogues. Hierarchical discourse structures in dialogues are explored in works like (Li et al. 2020; Zhu et al. 2020; Peng et al. 2022). The significance of speaker roles, providing insights into emotional states, is discussed in (Majumder et al. 2019; He et al. 2021). Efforts integrating external knowledge are seen in (Ren et al. 2020; Zhan et al. 2021; Li et al. 2023b), while syntactic parsing techniques are explored by (Fei et al. 2022b; Chen and Miyao 2022). Graph-based models, useful for tasks like emotion recognition and dialogue generation, are presented in (Shen et al. 2021; Hu et al. 2021; Chen et al. 2020; Feng et al. 2022; Lin et al. 2021; Liang et al. 2021). Inspired by these studies, we introduce a token-level heterogeneous graph to capture the intricacies of speaker roles and reply relationships, enriching dialogue feature understanding.

Preliminary

Task Definition

The DiaASQ task takes as its input a dialogue $D = \{(s_1, u_1), (s_2, u_2), \cdots, (s_{|D|}, u_{|D|})\}$ and a corresponding reply list $L = \{l_1, l_2, \cdots, l_{|D|}\}$. Here, $u_i = \{w_{i1}, w_{i2}, \cdots\}$ denotes the i-th utterance with w_{ij} representing its tokens, while s_i signifies the speaker of u_i. The list L indicates that the i-th utterance is a reply to the l_i-th utterance. The primary objective of this task is to extract a collection of quadruples $C = \{(t_i, a_i, o_i, p_i)\}_{i=1}^{|C|}$, where t_i, a_i, o_i and p_i are spans corresponds to the spans of the target, aspect, opinion, and polarity, respectively.

Label Schema

We transform quadruple extraction into token-pair relation detection by constructing an entity matrix and a relation matrix, from which we can further decode the quadruples by interpreting the labels of each token pair.

As illustrated in Figure 2, for the entity matrix, we define four labels: *target*, *aspect*, *opinion*, and *empty*. The *target* label identifies tokens from the head token to the tail token within a target term, while the *aspect* and *opinion* labels serve similar purposes for their respective items. The *empty* label is assigned when none of the three previously mentioned labels applies to the current token pair.

The relation matrix illustrates the relationships between head tokens of two items and encompasses five labels. The *rel* label describes relationships from one head token to another, in pairs such as (*target*, *aspect*) or (*aspect*, *opinion*). The labels *rel-pos*, *rel-neg*, and *rel-other* describe the relationships from the head of a *target* to the head of an *opinion*, representing positive, negative, and other sentiments, respectively. The *empty* label in this matrix indicates that no valid label has been assigned to the token pair.

To derive the quadruples, we adhere to the following decoding procedures:

- We first extract entity candidates categorized as *target*, *aspect*, and *opinion* from the entity matrix.
- We next recognize valid pairs by cross-referencing the relation matrix with the identified entity candidates.

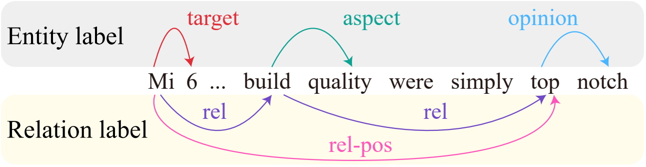


Figure 2: Label schema for entity matrix and relation matrix.

- Then, for each combination of (*target*, *aspect*, *opinion*): if all pairs within this combination are validated, they form a triplet. Together with the sentiment polarity between *target* and *opinion*, these components together constitute a quadruple.

Heterogeneous Graph

We construct a reply-speaker heterogeneous graph (R-S graph) to obtain token representations integrated with discourse features. In detail, we create a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes, i.e., the tokens in the dialogue utterance, and $\mathcal{E}$ is the set of edges with three categories, i.e., *reply*, *same speaker*, and *self-connection*. The first two edges are inherent features of discourse, while the self-connection is introduced to enhance model stability. Take *reply* relation as an example, for a token $v_i \in u_a$ and token $v_j \in u_b$, if u_a replies to u_b, we add an edge from v_i to v_j in the R-S graph.

Next, given a node v_i, we specify five meta-paths that contain multiple edges to control the information flow:

$$f(v_i) = \begin{cases} \text{Rep}: & v_j \xrightarrow{rep} v_i \\ \text{Spk}: & v_j \xrightarrow{spk} v_i \\ \text{Spk-rep}: & v_j \xrightarrow{rep} v_x \xrightarrow{spk} v_i \\ \text{Rep-spk}: & v_j \xrightarrow{rep} v_x \xrightarrow{spk} v_i \\ \text{Self}: & v_i \xrightarrow{self} v_i \end{cases} \quad (1)$$

Here v_x denotes an intermediate node. For each meta-path, the start node is defined as the neighbor of the target node v_i. For instance, in Figure 1, the path from the token 'battery' in u_4 to 'Xiaomi' in u_1 follows the Rep-spk meta-path. In this case, 'Xiaomi' is the target node, and 'battery' is one of its neighbors involved in the Rep-spk path.

Method

In this section, we delve into the architecture and functioning of our H2DT model. As illustrated in Figure 3, the model encompasses four components: text encoder, Reply-Speaker (R-S) Graph, entity extraction, and pair extraction.

Text Encoder

Considering the superiority demonstrated by pre-trained language models (PLMs) in the NLP domain, we employ a PLM to achieve deep contextualized token representations. Given a dialogue D, we flatten it and concatenate its utterances into a continuous sequence, using the specific tag [CLS] as a separator. This sequence is then fed into the PLM

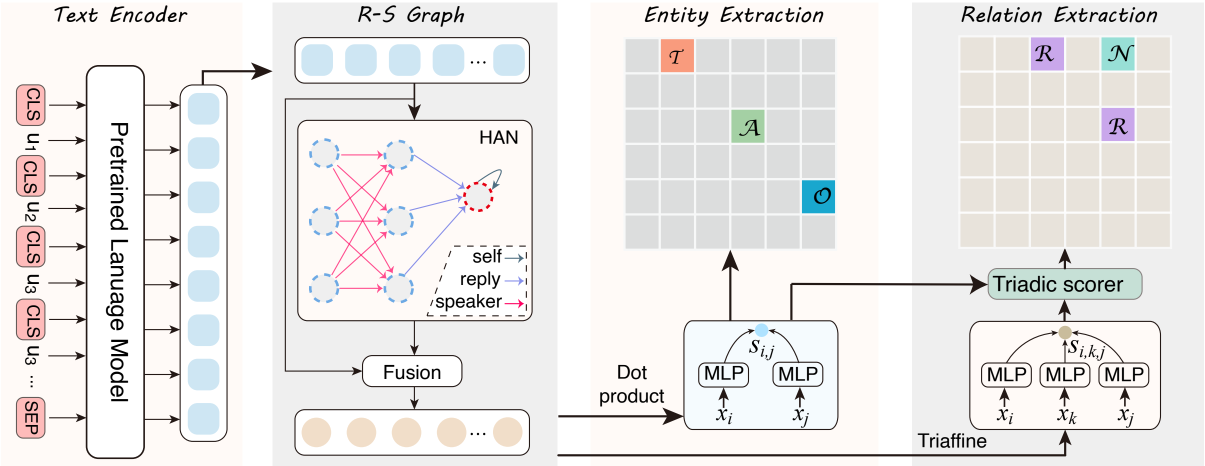


Figure 3: The overarching architecture of our H2DT model. Here, $\mathcal{T}$, $\mathcal{A}$, and $\mathcal{O}$ stand for *target*, *aspect*, and *opinion*, respectively. Similarly, $\mathcal{R}$ and $\mathcal{N}$ correspond to *relation* and *relation-negative*, respectively. Initially, a pre-trained language model is employed to procure contextualized representations for the entire dialogue. Subsequent to this, a heterogeneous attention network is applied to the reply-speaker graph. Then, we employ a dot product to formulate the entity matrix and, in further steps, capitalize on triadic interactions for token relation extraction.

to generate the desired token representation:

$$\text{Input} = \{[\text{CLS}], u_1, [\text{CLS}], u_2, \cdots, u_{|D|}, [\text{SEP}]\}, \quad (2)$$

$$[\boldsymbol{h}_0, \boldsymbol{h}_1, \cdots, \boldsymbol{h}_n] = \text{PLM}(\text{Input}), \quad (3)$$

where $[\boldsymbol{h}_0, \boldsymbol{h}_1, \cdots, \boldsymbol{h}_n]$ denotes all the utterance tokens in the dialogue, and $\boldsymbol{h}_i$ represents the representation of the i-th token. Note that the total length of each dialogue does not exceed the maximum input length permitted by the PLMs.

R-S Graph

Upon obtaining the representation for each token, we utilize the heterogeneous attention network (HAN, Wang et al. (2019)) to enrich the token representation with additional discourse features. Specifically, drawing upon the R-S graph, detailed in Section , we perform information integration for a token v_i with respect to meta-path Ω:

$$\boldsymbol{h}_i^{\Omega} = \sigma\big(\sum_{j \in V_i^{\Omega}} \alpha_{ij} \cdot \boldsymbol{h}_j\big), \quad (4)$$

$$\alpha_{ij} = \frac{\exp(\boldsymbol{W}_g[\boldsymbol{h}_i; \boldsymbol{h}_j])}{\sum_{k \in V_i^{\Omega}} \exp\left(\boldsymbol{W}_g[\boldsymbol{h}_k; \boldsymbol{h}_j]\right)}, \quad (5)$$

where $\boldsymbol{h}_i^{\Omega}$ denotes the representation of token v_i for meta-path Ω, σ is the sigmoid activation function, α_{ij} represents the attention weight, and V_i^{Ω} defines the set of neighbor nodes for token v_i concerning Ω.

Subsequently, we employ an additional attention block to integrate these representations from various meta-paths:

$$\gamma^{\Omega} = \frac{\exp(\text{att}_g(\boldsymbol{h}_i^{\Omega}))}{\sum_{\Omega} \exp(\text{att}_g(\boldsymbol{h}_i^{\Omega}))}, \quad (6)$$

$$\boldsymbol{h}_i^{g} = \sum_{\Omega} \gamma^{\Omega} \cdot \boldsymbol{h}_i^{\Omega}. \quad (7)$$

Here, att_g is a meta-path sensitive attention function, similar to the one described in Eq. (5). Furthermore, $\boldsymbol{h}_i^{g}$ represents the token v_i after applying HAN.

To retain the intrinsic information offered by the PLM, we introduce a fuse gate to dynamically combine the basic representation $\boldsymbol{h}_i$ with that derived from HAN:

$$F = \sigma(\boldsymbol{W}_f(\boldsymbol{h}_i^{g}, \boldsymbol{h}_i) + \boldsymbol{b}_g)\ , \quad (8)$$

$$\boldsymbol{x}_i = (1{-}F) * \boldsymbol{h}_i^{g} + F * \boldsymbol{h}_i\ , \quad (9)$$

where $\boldsymbol{x}_i$ represents the final representation of the token v_i.

Entity Extraction

Then we conduct entity label prediction. We use a label-wise MLP to transform the token representation:

$$\boldsymbol{k}_i^c = \text{MLP}_c^k(\boldsymbol{x}_i);\ \boldsymbol{m}_i^c = \text{MLP}_c^m(\boldsymbol{x}_i)\ , \quad (10)$$

where $c \in \{t, a, o, e\}$ denotes the label in token pair. Since the relative distance plays a vital role in token-pair label decision, we further incorporate RoPE (Su et al. 2021) to enhance the relative distance encoding.

$$\hat{\boldsymbol{k}}_i^c = \boldsymbol{\Phi}(\theta; i)\boldsymbol{k}_i^c;\ \hat{\boldsymbol{m}}_i^c = \boldsymbol{\Phi}(\theta; i)\boldsymbol{m}_i^c\ , \quad (11)$$

where $\Phi(\theta; i)$ is a transformation with parameter θ and index i, and $\hat{\boldsymbol{k}}_i^c \in \mathbb{R}^d$ has the same dimension with $\boldsymbol{k}_i^c$.

Then, we conduct dot-product for each token pair to obtain the label-wise score:

$$s_{i,j}^c = (\hat{\boldsymbol{k}}_i^c)^{\top} \hat{\boldsymbol{m}}_j^c\ , \quad (12)$$

where $s_{i,j}^c$ is the score of label between token v_i and token v_j in entity matrix. Finally, we can obtain the probability for the entity matrix label:

$$\boldsymbol{p}_{i,j}^{ent} = \text{Softmax}([s_{i,j}^t; s_{i,j}^a; s_{i,j}^o; s_{i,j}^e])\ , \quad (13)$$

where $\boldsymbol{p}_{i,j}^{ent}$ denotes the probability distribution for the pair (v_i, v_j) in the entity matrix.

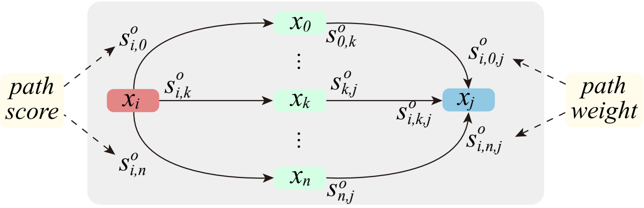


Figure 4: Details about triadic scorer. $s^o_{i,k}$ denote the path score from s^o_i to s^o_k, and $s^o_{i,k,j}$ denotes the traffine score, servers as weight for path $i \to k \to j$.

Relation Extraction

For the prediction of the relation matrix, we initially compute the pair-wise score, mirroring the techniques laid out in Eq. (10) through Eq. (12). This score is denoted as $s^o_{i,j}$, signifying the unary score for the token pair (v_i, v_j). The label o is chosen from the set {*rel*, *rel-pos*, *rel-neg*, *rel-other*, *empty*}. Following this, we employ the triaffine operation (Zhang, Li, and Zhang 2020) to determine the score for each triplet (v_i, v_k, v_j):

$$\boldsymbol{z}^o_i, \boldsymbol{z}^o_k, \boldsymbol{z}^o_j = \text{MLP}^o_1(\boldsymbol{x}_i), \text{MLP}^o_2(\boldsymbol{x}_k), \text{MLP}^o_3(\boldsymbol{x}_j) \ , \tag{14}$$

$$\hat{s}^o_{i,k,j} = \boldsymbol{W}^o_t \begin{bmatrix} \boldsymbol{z}^o_i \\ 1 \end{bmatrix} \boldsymbol{z}^o_k \boldsymbol{z}^o_j \ , \tag{15}$$

$$s^o_{i,k,j} = \frac{\exp(\hat{s}^o_{i,k,j})}{\sum_o \exp(\hat{s}^o_{i,k,j})} \ , \tag{16}$$

where $s^o_{i,k,j}$ represents the unary score for the triplet (v_i, v_k, v_j) associated with label o, $\boldsymbol{W}^o_t \in \mathbb{R}^{d' \times d' \times (d'+1)}$ is the parameter, and d' is the dimension of z^o_*. This triaffine operation can be efficiently implemented using the einsum function in PyTorch.

To compute the score for each pair (v_i, v_j), each third token is treated as a bridge. The triaffine score $s^o_{i,k,j}$ is employed as the weight measure for this computation (Zhou et al. 2022). Using this, all path scores are fused to yield the final score for the pair (v_i, v_j):

$$q^o_{i,j} = s^o_{i,j} + \sum_k (s^o_{i,k} + s^o_{k,j}) * s^o_{i,k,j} \ , \tag{17}$$

where $q^o_{i,j}$ is the conclusive score between tokens v_i and v_j, incorporating all third token scores. As shown in Figure 4, each $s^o_{i,k,j}$ serves as a weight for fusing the path score for every path that passes from v_i to v_j. By implementing this procedure for each label, predictions regarding the relationship label can be ascertained:

$$\boldsymbol{p}^{rel}_{i,j} = \text{Softmax}([q^r_{i,j}; q^{r\text{-}p}_{i,j}; q^{r\text{-}n}_{i,j}; q^{r\text{-}o}_{i,j}; q^e_{i,j}]) \ , \tag{18}$$

where $\boldsymbol{p}^{rel}_{i,j}$ denotes the probability distribution of the relation label for the pair (v_i, v_j).

Learning Objectives

To refine the performance of our model, we undertake a joint optimization process of the loss functions associated with both the entity and relation matrices:

$$\mathcal{L}_{ent} = -\frac{1}{n^2} \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{n} \log \boldsymbol{p}^{ent}_{i,j}[y^{ent}_{i,j}] \ , \tag{19}$$

$$\mathcal{L}_{rel} = -\frac{1}{n^2} \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{n} \log \boldsymbol{p}^{rel}_{i,j}[y^{rel}_{i,j}] \ , \tag{20}$$

$$\mathcal{L} = \mathcal{L}_{ent} + \mathcal{L}_{rel} \ , \tag{21}$$

where n represents the token count in the dialogue. The indices $y^{ent}_{i,j}$ and $y^{rel}_{i,j}$ represent the gold standard labels for the entity and relation matrices. The loss functions for the entity and relation matrices are denoted by $\mathcal{L}_{ent}$ and $\mathcal{L}_{rel}$, with $\mathcal{L}$ signifying the total loss being optimized during training.

In the inference stage, the model predicts labels for the entity and relation matrices. Subsequently, the methodology described in Section is applied for quadruple decoding.

Experiment

Experimental Settings

Dataset We rigorously conducted experiments on the DiaASQ dataset (Li et al. 2023a), which contains both Chinese and English versions. This corpus consists of multipart, multi-turn dialogues sourced from social media, predominantly focusing on topics related to mobile phones.

Baselines For a comprehensive comparison, we incorporated the following models as our baselines: **CRF-Extract** (Cai, Xia, and Yu 2021), **SpERT** (Eberts and Ulges 2020), **ParaPhrase** (Zhang et al. 2021a), **Span-ASTE** (Xu, Chia, and Bing 2021), and **Meta-WP** (Li et al. 2023a).

Evaluation Following previous work (Li et al. 2023a), we assess our experiments using precision, recall, and F1 score metrics. These metrics are employed for item detection ($\mathcal{T}$, $\mathcal{A}$, $\mathcal{O}$), pair detection ($\mathcal{T}$-$\mathcal{A}$, $\mathcal{T}$-$\mathcal{O}$, $\mathcal{A}$-$\mathcal{O}$), triplet detection ($\mathcal{T}$-$\mathcal{A}$-$\mathcal{O}$), and quadruple detection (the full quadruple). An exact match with the gold standard is required for an item to be considered a correct prediction.

Hyperparameters Consistent with prior research, we initialize our pre-trained language model (PLM) with Chinese-Roberta-wwm-ext (Cui et al. 2021) and Roberta-Large (Liu et al. 2019) for Chinese and English dataset, respectively. The hidden dimensions for the HAN layer are set to 768 and 1024, aligning with the PLM output dimensions. For the MLP in Eq. (10), the dimension is set to 256, respectively. The dimension in Eq. (14) is set to 100 and 50 for analysis. During training, we employ a batch size of 2 and run for 15 epochs. The optimizer is AdamW with a learning rate of 1e-5 and weight decay of 1e-8. We report the average result for each experiment with four varying seed values to mitigate random factor influences.

Main Results

Table 1 showcases the primary results of our experiments.

Item Detection: We observe that the H2DT model provides only modest improvements for the item detection task on both datasets. This is due to the relative maturity of the task of detecting target, aspect, and opinion items. As it does not involve complex interactions, even naive models can achieve

Data	Methods	Entity (F1)			Pair(F1)			Triplet			Quadruple		
		$\mathcal{T}$	$\mathcal{A}$	$\mathcal{O}$	$\mathcal{T}$-$\mathcal{A}$	$\mathcal{T}$-$\mathcal{O}$	$\mathcal{A}$-$\mathcal{O}$	P	R	F	P	R	F
ZH	CRF-Extract	91.11	75.24	50.06	32.47	26.78	18.90	/	/	9.25	/	/	8.81
	SpERT	90.69	76.81	54.06	38.05	31.28	21.89	/	/	14.19	/	/	13.00
	ParaPhrase	/	/	/	37.81	34.32	27.76	/	/	27.98	/	/	23.27
	Span-ASTE	/	/	/	44.13	34.46	32.21	/	/	30.85	/	/	27.42
	Meta-WP	90.23	**76.94**	59.35	48.61	43.31	45.44	/	/	37.51	/	/	34.94
	H2DT	**91.72**	76.93	**61.87**	**50.48**	**48.80**	**52.40**	**45.40**	**40.50**	**42.81**	**42.78**	**38.17**	**40.34**
	Δ	+0.61	-0.01	+2.52	+1.87	+5.49	+6.96	/	/	+5.30	/	/	+5.40
EN	CRF-Extract	88.31	71.71	47.90	34.31	20.94	19.21	/	/	12.80	/	/	11.59
	SpERT	87.82	74.65	54.17	28.33	21.39	23.64	/	/	13.38	/	/	13.07
	ParaPhrase	/	/	/	37.22	32.19	30.78	/	/	26.76	/	/	24.54
	Span-ASTE	/	/	/	42.19	30.44	45.90	/	/	28.34	/	/	26.99
	w/o PLM	/	/	/	27.26	20.63	44.62	/	/	14.17	/	/	13.84
	Meta-WP	88.62	**74.71**	60.22	47.91	45.58	44.27	/	/	36.80	/	/	33.31
	H2DT	**88.69**	73.81	**62.61**	**48.69**	**48.84**	**52.47**	**44.36**	**40.23**	**42.19**	**41.01**	**37.20**	**39.01**
	Δ	+0.07	-0.90	+2.39	+0.78	+3.26	+6.57	/	/	+5.39	/	/	+5.70

Table 1: Main results on DiaASQ dataset. 'ZH' and 'EN' denote the Chinese and English datasets, respectively. The number with bold is the best result, and that with waveline denotes the second best result.

high performance. **Pair Detection:** Our model exhibits average performance for $\mathcal{T}$-$\mathcal{A}$ pair detection. Nevertheless, a marked improvement is evident for the $\mathcal{T}$-$\mathcal{O}$ pair, with gains of 5.49 and 3.26 in F1 scores on the Chinese and English datasets, respectively. The $\mathcal{A}$-$\mathcal{O}$ pair detection showcases improvements of 6.96 and 6.57 in F1 scores. The results show that the triadic interaction mechanism significantly improves $\mathcal{T}$-$\mathcal{O}$ and $\mathcal{A}$-$\mathcal{O}$ detections.

Triplet Extraction: Noteworthy improvements in triplet extraction are evident with our model, achieving F1 scores of 5.30 and 5.29 on Chinese and English datasets, respectively. These figures clearly demonstrate the effectiveness and robustness of our approach. **Quadruple Extraction:** Focusing on quadruple extraction, a core task of DiaASQ, we notice marked improvements. The data shows gains of 5.40 and 5.70 in F1 scores on Chinese and English datasets, respectively. These findings underscore the superiority of our H2DT model for DiaASQ over existing baselines.

In conclusion, the results from our experiments provide compelling evidence of the strength and stability of the H2DT model, especially in tasks involving intricate interactions such as triplet and quadruple extractions. The consistent improvements across multiple tasks underline the potential of the model to become a strong benchmark in DiaASQ.

Ablation Study

To rigorously investigate the contributions of each module, we conducted an ablation study, as shown in Table 2. Firstly, concerning the R-S graph, we observed that its removal led to a notable decline in performance. The F1 scores for the quadruple dropped by approximately 3.32 and 1.83 for the Chinese and English datasets, respectively. Subsequently, on examining different modules in the R-S graph, the performance demonstrated varying degrees of decrement upon their removal. The 'self-link' exhibited the least decline, while 'First' and 'Second' experienced more substantial drops. This suggests that our R-S graph is adept at capturing conversational features, which in turn influences the extraction of triplets and quadruples. Furthermore, when comparing the results after removing the 'triadic' module, the decrease in performance was the most pronounced, reaching up to 2.71 and 2.61 in F1 scores on the Chinese and English datasets, respectively. This underscores the importance of our triadic interaction scores for the extraction process.

Methods	**Chinese(F1)**		**English(F1)**	
	Trip.	**Quad.**	**Trip.**	**Quad.**
H2DT	42.81	40.34	42.19	39.01
w/o R-S Graph	40.99	37.02	40.66	37.18
w/o Self	42.42	39.10	41.66	38.86
w/o First	41.49	38.70	41.71	38.44
w/o Second	42.08	38.97	41.13	38.30
w/o Triadic	40.51	37.63	39.58	36.40

Table 2: Ablation study focusing on the R-S graph and triadic interaction. 'Self' represents the self-link in the R-S graph, while 'first' refers to the first-order link, i.e., Rep and Spk. 'Second' signifies the second-order link, encompassing Rep-spk and Spk-rep. Further details can be found in Eq. (1). The term 'w/o triadic' indicates the removal of the second part in Eq. (17).

In-depth Analysis

Subsequently, we conduct a detailed analysis to address several questions to further validate the efficacy of our model.

Q1: How does the R-S Graph enhance extraction performance? As demonstrated in Figure 5, we undertook a detailed comparative analysis to investigate the quadruple ex-

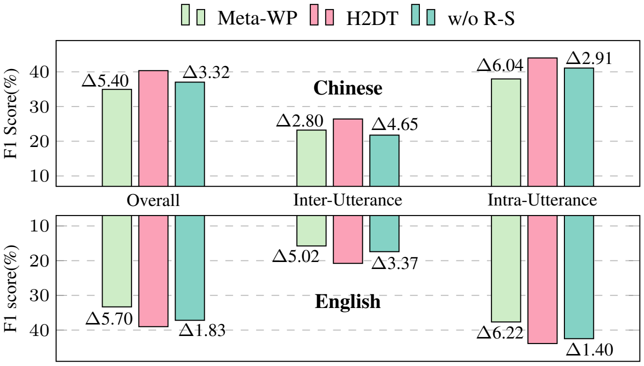


Figure 5: Quadruple extraction scores on overall, inter-utterance, and intra-utterance instances. The term 'w/o R-S graph' denotes the removal of the R-S graph in our H2DT model, and Δx indicates the performance gap compared to the H2DT model.

traction capabilities of three distinct models: the Meta-WP model, the H2DT model, and a version of the H2DT model from which the R-S Graph component has been excluded. For this comparative study, we categorized the results into three primary dimensions: the overall performance metrics, metrics for intra-utterance quadruples, and those for inter-utterance quadruples. Upon integrating the R-S Graph, our model exhibits a significant improvement in the extraction of the inter-utterance quadruples. Specifically, considerable increments of 4.65 and 3.37 are observed in the F1 scores for the Chinese and English datasets, respectively. This observation substantiates that the proposed R-S Graph can effectively harness discourse features, thereby elevating the extraction of inter-utterance quadruples. Furthermore, it can be noted that the extraction of intra-utterance quadruples is somewhat influenced by the context. As such, our proposed R-S Graph demonstrates a marginal yet discernible impact on improving the extraction of intra-utterance quadruples. Consequently, the average performance lies between the two aforementioned categories, with improvements of 3.32 and 1.83 in F1 scores for Chinese and English datasets, respectively. Additionally, the performances of H2DT on intra-utterance or inter-utterance quadruple extraction both notably surpass the Meta-WP model, underlining the superiority and robustness of H2DT.

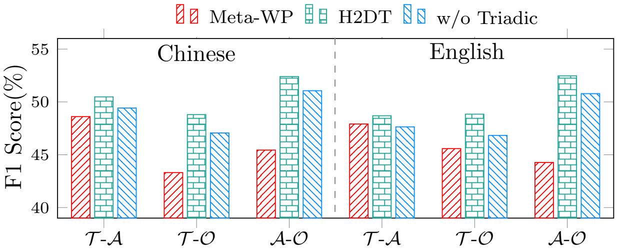


Figure 6: Performance of various pair extractions. The term 'w/o Triadic' denotes exclusion of Eq. (17)'s second part.

Methods	**Param.(M)**		**Speed(Dia./s)**	
	ZH	**EN**	**ZH**	**EN**
Meta-WP	114.05	363.91	5.33	3.66
H2DT	111.81	359.04	8.16	6.01
w/o Triadic	110.85	358.91	9.43	6.47

Table 3: Comparison of parameter counts and training speeds between H2DT and Meta-WP. 'Param.' represents the total number of parameters, while 'Dia./s' indicates the number of dialogues processed per second.

Q2: What impact does the triadic scorer have on different types of Pair combinations? We conduct an analysis encompassing diverse pair combinations with and without the incorporation of the triadic scorer. The results shown in Figure 6 indicate that the augmentation achieved through the utilization of the triadic scorer yielded relatively modest improvements in the case of the $\mathcal{T}$-$\mathcal{A}$ combination. A possible reason for this lies in the fact that both $\mathcal{T}$ and $\mathcal{A}$ are common terms describing an aspect of a target, making their combination relatively straightforward. However, when addressing $\mathcal{T}$-$\mathcal{O}$/$\mathcal{A}$-$\mathcal{O}$ combinations, especially $\mathcal{T}$-$\mathcal{O}$, the connection is weaker. Notably, the introduction of the triadic scorer brought about a significant improvement in these instances. This suggests that the utilization of intermediary tokens effectively addresses the fragmentation issue of pairs in relation to triplets, enhancing the coherence of the triplets and subsequently improving extraction performances.

Model Efficiency Analysis

A potential concern might be that our model structure, which facilitates extensive token interactions, could lead to an increase in the number of parameters. To address this apprehension, we enumerated both the parameter count and the training time. As shown in Table 3, it becomes evident that our model not only did not increase the parameter count but on the contrary, substantially reduced it compared to Meta-WP, while also enhancing the inference speed. Moreover, the introduction of the triadic component only contributed to a negligible increase in parameters (less than 1%). This underscores the efficiency of our proposed model.

Conclusion

In this paper, we proposed a model leveraging unified discourse features and triadic interaction for dialogue sentiment quadruple extraction. In detail, we build a heterogeneous graph to unify the encoding of the reply and speaker information in dialogue, achieving a comprehensive discourse structure. By harnessing the power of our triadic scorer, we further enhance the coherence and cohesion within the quadruple structure, ensuring robust token relationships. The experiment on benchmark demonstrates that our model achieves unquestionable leading performance on both two datasets of the DiaASQ task. We believe our work lays a solid foundation for future research in aspect-based sentiment analysis in dialogues.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 62176187), CCF-Baidu Open Fund, and China Scholarship Council (CSC).

References

Cai, H.; Xia, R.; and Yu, J. 2021. Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions. In *Proceedings of ACL*, 340–350.

Chen, B.; and Miyao, Y. 2022. Syntactic and Semantic Uniformity for Semantic Parsing and Task-Oriented Dialogue Systems. In *Findings of EMNLP*, 855–867.

Chen, L.; Lv, B.; Wang, C.; Zhu, S.; Tan, B.; and Yu, K. 2020. Schema-Guided Multi-Domain Dialogue State Tracking with Graph Attention Neural Networks. In *AAAI*, 7521–7528.

Chen, S.; Wang, Y.; Liu, J.; and Wang, Y. 2021. Bidirectional Machine Reading Comprehension for Aspect Sentiment Triplet Extraction. In *AAAI*, 12666–12674.

Cui, Y.; Che, W.; Liu, T.; Qin, B.; and Yang, Z. 2021. Pre-Training With Whole Word Masking for Chinese BERT. *IEEE ACM Trans. Audio Speech Lang. Process.*, 29: 3504–3514.

Eberts, M.; and Ulges, A. 2020. Span-Based Joint Entity and Relation Extraction with Transformer Pre-Training. In *Proceedings of ECAI*, volume 325, 2006–2013.

Fan, Z.; Wu, Z.; Dai, X.-Y.; Huang, S.; and Chen, J. 2019. Target-oriented Opinion Words Extraction with Target-fused Neural Sequence Labeling. In *Proceedings of the NAACL*, 2509–2518.

Fei, H.; Li, F.; Li, C.; Wu, S.; Li, J.; and Ji, D. 2022a. Inheriting the Wisdom of Predecessors: A Multiplex Cascade Framework for Unified Aspect-based Sentiment Analysis. In *IJCAI*, 4121–4128.

Fei, H.; Li, J.; Wu, S.; Li, C.; Ji, D.; and Li, F. 2022b. Global Inference with Explicit Syntactic and Discourse Structures for Dialogue-Level Relation Extraction. In *IJCAI*, 4107–4113.

Fei, H.; Zhang, Y.; Ren, Y.; and Ji, D. 2020. Latent Emotion Memory for Multi-Label Emotion Classification. In *AAAI*, 7692–7699.

Feng, Y.; Lipani, A.; Ye, F.; Zhang, Q.; and Yilmaz, E. 2022. Dynamic Schema Graph Fusion Network for Multi-Domain Dialogue State Tracking. In *Proceedings of ACL*, 115–126.

Gao, L.; Wang, Y.; Liu, T.; Wang, J.; Zhang, L.; and Liao, J. 2021. Question-Driven Span Labeling Model for Aspect-Opinion Pair Extraction. In *AAAI*, 12875–12883.

He, Z.; Tavabi, L.; Lerman, K.; and Soleymani, M. 2021. Speaker Turn Modeling for Dialogue Act Classification. In *Findings of the EMNLP*, 2150–2157.

Hu, J.; Liu, Y.; Zhao, J.; and Jin, Q. 2021. MMGCN: Multimodal Fusion via Deep Graph Convolution Network for Emotion Recognition in Conversation. In *Proceedings of ACL*, 5666–5675.

Li, B.; Fei, H.; Li, F.; Wu, Y.; Zhang, J.; Wu, S.; Li, J.; Liu, Y.; Liao, L.; Chua, T.-S.; and Ji, D. 2023a. DiaASQ: A Benchmark of Conversational Aspect-based Sentiment Quadruple Analysis. In *Findings of ACL*, 13449–13467.

Li, J.; Ji, D.; Li, F.; Zhang, M.; and Liu, Y. 2020. HiTrans: A Transformer-Based Context- and Speaker-Sensitive Model for Emotion Detection in Conversations. In *Proceedings of the COLING*, 4190–4200.

Li, W.; Zhu, L.; Mao, R.; and Cambria, E. 2023b. SKIER: A Symbolic Knowledge Integrated Model for Conversational Emotion Recognition. In *AAAI*, 13121–13129.

Liang, Y.; Meng, F.; Zhang, Y.; Chen, Y.; Xu, J.; and Zhou, J. 2021. Infusing Multi-Source Knowledge with Heterogeneous Graph Neural Network for Emotional Conversation Generation. In *AAAI*, 13343–13352.

Lin, S.; Zhou, P.; Liang, X.; Tang, J.; Zhao, R.; Chen, Z.; and Lin, L. 2021. Graph-Evolving Meta-Learning for Low-Resource Medical Dialogue Generation. In *AAAI*, 13362–13370.

Liu, P.; Joty, S.; and Meng, H. 2015. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. In *Proceedings of EMNLP*, 1433–1443.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. *CoRR*, abs/1907.11692.

Majumder, N.; Poria, S.; Hazarika, D.; Mihalcea, R.; Gelbukh, A. F.; and Cambria, E. 2019. DialogueRNN: An Attentive RNN for Emotion Detection in Conversations. In *AAAI*, 6818–6825.

Mao, Y.; Shen, Y.; Yu, C.; and Cai, L. 2021. A Joint Training Dual-MRC Framework for Aspect Based Sentiment Analysis. In *AAAI*, 13543–13551.

Peng, H.; Xu, L.; Bing, L.; Huang, F.; Lu, W.; and Si, L. 2020. Knowing What, How and Why: A Near Complete Solution for Aspect-Based Sentiment Analysis. In *AAAI*, 8600–8607.

Peng, W.; Hu, Y.; Xing, L.; Xie, Y.; Sun, Y.; and Li, Y. 2022. Control Globally, Understand Locally: A Global-to-Local Hierarchical Graph Network for Emotional Support Conversation. In *IJCAI*, 4324–4330.

Peper, J.; and Wang, L. 2022. Generative Aspect-Based Sentiment Analysis with Contrastive Learning and Expressive Structure. In *Findings of the EMNLP*, 6089–6095.

Picard, R. W. 2000. *Affective computing*. MIT press.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manandhar, S.; AL-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; Hoste, V.; Apidianaki, M.; Tannier, X.; Loukachevitch, N.; Kotelnikov, E.; Bel, N.; Jiménez-Zafra, S. M.; and Eryiğit, G. 2016. SemEval-2016 Task 5: Aspect Based Sentiment Analysis. In *Proceedings of the SemEval-2016*, 19–30.

Pontiki, M.; Galanis, D.; Papageorgiou, H.; Manandhar, S.; and Androutsopoulos, I. 2015. SemEval-2015 Task 12: Aspect Based Sentiment Analysis. In *Proceedings of the SemEval 2015*, 486–495.

Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.; Androutsopoulos, I.; and Manandhar, S. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the (SemEval 2014)*, 27–35.

Ren, P.; Chen, Z.; Monz, C.; Ma, J.; and de Rijke, M. 2020. Thinking Globally, Acting Locally: Distantly Supervised Global-to-Local Knowledge Selection for Background Based Conversation. In *AAAI*, 8697–8704.

Shen, W.; Wu, S.; Yang, Y.; and Quan, X. 2021. Directed Acyclic Graph Network for Conversational Emotion Recognition. In *Proceedings of ACL*, 1551–1560.

Su, J.; Lu, Y.; Pan, S.; Wen, B.; and Liu, Y. 2021. RoFormer: Enhanced Transformer with Rotary Position Embedding. *CoRR*, abs/2104.09864.

Wang, W.; Pan, S. J.; Dahlmeier, D.; and Xiao, X. 2016. Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis. In *Proceedings of EMNLP*, 616–626.

Wang, X.; Ji, H.; Shi, C.; Wang, B.; Ye, Y.; Cui, P.; and Yu, P. S. 2019. Heterogeneous Graph Attention Network. In *Proceedings of WWW*, 2022–2032.

Wei, P.; Xu, N.; and Mao, W. 2019. Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity. In *Proceedings of EMNLP*, 4787–4798.

Wu, S.; Fei, H.; Ren, Y.; Ji, D.; and Li, J. 2021. Learn from Syntax: Improving Pair-wise Aspect and Opinion Terms Extraction with Rich Syntactic Knowledge. In *IJCAI*, 3957–3963.

Xu, L.; Chia, Y. K.; and Bing, L. 2021. Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction. In *Proceedings of ACL*, 4755–4766.

Yan, H.; Dai, J.; Ji, T.; Qiu, X.; and Zhang, Z. 2021. A Unified Generative Framework for Aspect-based Sentiment Analysis. In *Proceedings of ACL*, 2416–2429.

Yang, Y.; Li, K.; Quan, X.; Shen, W.; and Su, Q. 2020. Constituency Lattice Encoding for Aspect Term Extraction. In *Proceedings of COLING*, 844–855.

Zhan, H.; Zhang, H.; Chen, H.; Ding, Z.; Bao, Y.; and Lan, Y. 2021. Augmenting Knowledge-grounded Conversations with Sequential Knowledge Transition. In *Proceedings of NAACL*, 5621–5630.

Zhang, M.; Zhu, Y.; Liu, Z.; Bao, Z.; Wu, Y.; Sun, X.; and Xu, L. 2023a. Span-level Aspect-based Sentiment Analysis via Table Filling. In *Proceedings of ACL*, 9273–9284.

Zhang, W.; Deng, Y.; Li, X.; Yuan, Y.; Bing, L.; and Lam, W. 2021a. Aspect Sentiment Quad Prediction as Paraphrase Generation. In *Proceedings of EMNLP*, 9209–9219.

Zhang, W.; Li, X.; Deng, Y.; Bing, L.; and Lam, W. 2021b. Towards Generative Aspect-Based Sentiment Analysis. In *Proceedings of ACL*, 504–510.

Zhang, W.; Li, X.; Deng, Y.; Bing, L.; and Lam, W. 2023b. A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges. *IEEE Trans. Knowl. Data Eng.*, 35(11): 11019–11038.

Zhang, Y.; Li, Z.; and Zhang, M. 2020. Efficient Second-Order TreeCRF for Neural Dependency Parsing. In *Proceedings of ACL*, 3295–3305.

Zhou, S.; Xia, Q.; Li, Z.; Zhang, Y.; Hong, Y.; and Zhang, M. 2022. Fast and Accurate End-to-End Span-based Semantic Role Labeling as Word-based Graph Parsing. In *Proceedings of COLING*, 4160–4171.

Zhu, H.; Nan, F.; Wang, Z.; Nallapati, R.; and Xiang, B. 2020. Who Did They Respond to? Conversation Structure Modeling Using Masked Hierarchical Transformer. In *AAAI*, 9741–9748.