Expert Systems With Applications 261 (2025) 125556

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

IFusionQuad: A novel framework for improved aspect-based sentiment quadruple analysis in dialogue contexts with advanced feature integration and contextual CloBlock

Haoyu Jiang [a], Xiaoliang Chen [a,b,d,*], Duoqian Miao [b], Hongyun Zhang [b], Xiaolin Qin [c], Xu Gu [c,a], Peng Lu [d]

[a] School of Computer and Software Engineering, Xihua University, Chengdu 610039, PR China
[b] College of Electronic and Information Engineering, Tongji University, Shanghai 201804, PR China
[c] Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041, PR China
[d] Department of Computer Science and Operations Research, University of Montreal, Montreal, QC H3C3J7, Canada

ARTICLE INFO

Dataset link: https://github.com/Joeisjoejoe/IFusionQuad

Keywords:
Natural language processing
Aspect-based sentiment analysis
Aspect sentiment quadruple extraction
DiaASQ

ABSTRACT

Aspect-based sentiment analysis (ABSA) represents a crucial field of natural language processing (NLP). It focuses on deriving detailed sentiment insights from textual content. Dialogue-level aspect-based sentiment quadruple extraction (DiaASQ) is specifically concerned with pinpointing target-aspect-opinion-emotion quadruples within conversations. DiaASQ is important in industries like e-commerce, social media analytics, and customer feedback. However, Current ABSA approaches predominantly focus on single-text scenarios, often overlooking the complexities involved in sentiment analysis within conversational contexts. To fill this gap, this paper presents the IFusionQuad model, which is specifically designed for the DiaASQ task. Our contributions include the innovative integration of CloBlock in ABSA, enhancing feature representation with context-aware weights. The InteractiveNet Fusion Module further advances dialogue understanding by aggregating dialogue-specific features such as threads, speakers, and replies. Components such as CloBlock, gating mechanism, and Biaffine attention effectively mitigate data noise issues, improving the relevance of feature extraction. Empirical evaluation on standard datasets demonstrates that the IFusionQuad model outperforms baseline methods, achieving substantial improvements in quadruple extraction. Specifically, our model shows a 6.59% increase in micro F1 and a 7.05% increase in identification F1 for Chinese datasets, and a 2.65% and 4.69% increase in micro F1 and identification F1, respectively, for English datasets. The results clearly demonstrate our IFusionQuad model's efficacy, which consistently outperforms baseline models across all evaluation datasets on the DiaASQ task.

1. Introduction

Sentiment analysis research seeks to equip machines with the ability to understand human perspectives and emotions, thereby enhancing human–computer interaction. Aspect-based sentiment analysis (ABSA) advances this field by focusing on identifying sentiment tied to specific aspects within a text, allowing for a more nuanced understanding of sentiment. However, current ABSA research is largely confined to single-text scenarios, such as sentences or documents. For example, the SemEval benchmarks (Pontiki et al., 2016), (Pontiki et al., 2015), (Pontiki et al., 2014), extensively utilized in ABSA research, provide only sentence-level annotations. This constraint limits the applicability of ABSA in dynamic and interactive scenarios. In real-world contexts, the scope of ABSA expands beyond static texts to encompass conversational settings. Social media platforms such as Instagram, TikTok, and YouTube foster discussions that unfold over multiple rounds and turns, involving several participants. These dialogic interactions offer a more intricate data source for sentiment analysis. Despite the evident relevance of dialogue-level ABSA, research in this domain remains in its infancy.

Existing ABSA approaches, while effective for static analysis, struggle with the complexities of dialogue-based sentiment analysis. These methods often fail to maintain context across multiple turns and speakers, leading to a loss of nuanced sentiment detection. Additionally, real-world dialogues frequently involve overlapping conversational threads

* Corresponding author at: School of Computer and Software Engineering, Xihua University, Chengdu 610039, PR China.
E-mail addresses: jianghaoyu@stu.xhu.edu.cn (H. Jiang), chenxl@mail.xhu.edu.cn (X. Chen), dqmiao@tongji.edu.cn (D. Miao), zhanghongyun@tongji.edu.cn (H. Zhang), qinxl@casit.com.cn (X. Qin), 212021081200018@stu.xhu.edu.cn (X. Gu), peng.lu@umontreal.ca (P. Lu).

https://doi.org/10.1016/j.eswa.2024.125556
Received 31 July 2024; Received in revised form 7 October 2024; Accepted 11 October 2024
Available online 18 October 2024

and cross-talk, complicating the sentiment analysis task. These challenges underscore the need for sophisticated models that can capture contextual dependencies across extended dialogue sequences. Enhancing the capability to analyse sentiments in conversational contexts has significant implications, including improved customer service interactions and better insights into public opinion on social media.

This study bridges this gap by focusing on Dialogue-level Aspect-based Sentiment Quadruple extraction (DiaASQ), a task first introduced by Li et al. (2023). The DiaASQ task aims to identify the quadruples in the dialogue. Each quadruple is composed of four essential components: aspect terms (aspect), opinion terms (opinion), sentiment polarity associated with aspect-opinion pairs, and the target of aspect-opinion pairs. Let us consider the example dialogue shown in Fig. 1:

< **Person 1** >: Hey, have you guys seen that new TikTok trend? It is everywhere!

< **Person 2 – 4** >:.....

< **Person 5** >: Yep! Short video is so funny.

From this dialogue, we can extract the following sentiment quadruples:

- (TikTok, trend, everywhere, positive)
- (Instagram, layout, organized, positive)
- (Instagram, experience, mature, positive)
- (TikTok, short video, attractive, positive)
- (TikTok, short video, so funny, positive)

In these examples, categories are highlighted in green, aspects in blue, opinions in red, and their associated sentiment polarities in purple. This process illustrates the depth and complexity of DiaASQ. The dialogue utterances, attributed to their respective speakers (indicated on the left), are organized into a framework of responses.

The DiaASQ dataset (Li et al., 2023), comprising 1000 dialogue segments totalling 7452 utterances, provides a robust foundation for this study. Analysis of this dataset reveals an average of five participants per conversation, with approximately 22.2% of the quadruples exhibiting cross-talk. These statistics highlight the complexity of extracting sentiment information from dialogues compared to single-text scenarios. The DiaASQ task presents two principal challenges. First, elements within the quadruples are distributed across conversations with intricate response structures. This complexity requires models to extract sentiment information based on a comprehensive understanding of the discourse, making quadruple extraction within conversations more difficult than within isolated sentences. Additionally, DiaASQ requires the extraction of three elements and the categorization of one element, thereby increasing the task's difficulty.

Popular end-to-end ABSA systems, such as graph-based approaches (Chen et al., 2022; Zhou et al., 2021) and sequence-to-sequence methods (Mukherjee et al., 2021; Zhang et al., 2021b), face significant hurdles when applied to the DiaASQ task.

Graph-based methods face challenges related to computational inefficiency, particularly when enumerating all possible combinations of target, aspect, and opinion terms in dialogue data. The complexity of constructing and maintaining these graphs increases exponentially with the length and diversity of dialogues. Additionally, these methods are less effective at capturing long-distance dependencies and nuanced conversational contexts that are crucial for dialogue understanding. The reliance on dependency parsing and graph construction also introduces noise, especially when dealing with informal language and incomplete sentences that are typical in dialogues. Sequence-to-sequence methods suffer from exposure bias, where the model is trained on true target sequences but relies on its own generated output during inference. This discrepancy can lead to cascading errors, with initial mistakes propagating through the generated sequence, resulting in inaccurate quadruple extraction. Furthermore, Seq2Seq models often struggle to capture long-range dependencies and integrate context across multiple turns of conversation, making them less effective for tasks requiring a nuanced understanding of dialogue flows. These methods also exhibit weaker generalization capabilities when trained on limited or domain-specific data, which is common in real-world applications. These limitations highlight the need for novel approaches that can efficiently process the complex structures and contextual dependencies of dialogue data.

To overcome these challenges, we propose an end-to-end model called IFusionQuad for DiaASQ extraction. This approach encodes each discourse independently using a pre-trained language model (PLM) such as BERT (Devlin et al., 2019). IFusionQuad introduces an InteractiveNet Fusion layer that captures multiple dialogue features, including threads, speaker identities, and replying relationships, thereby reducing computational complexity compared to graph-based methods while enhancing the model's capability to handle long-distance dependencies. Drawing inspiration from the work of Fan et al. (2023), the CloBlock module further improves feature extraction by using convolution and pooling operations, followed by a gating mechanism to filter out noise, thereby mitigating the effects of exposure bias typical in Seq2Seq methods. Biaffine attention is incorporated to direct focus on crucial conversational features, allowing for accurate extraction of targets, aspects, and opinions. This creates a robust and scalable solution for DiaASQ tasks, allowing efficient and precise quadruple extraction in various dialogue settings.

Due to PLM constraints, discourse encoding is often limited to individual instances, which can harm conversational coherence. To address conversational coherence, the integration of Rotational Positional Embedding (RoPE) is proposed to capture relative discourse distances dynamically, enhancing the model's comprehension of discourse nuances.

Our contributions are succinctly summarized as follows:

- **Advancing ABSA with CloBlock.** We present the CloBlock module, a novel architectural innovation for ABSA that integrates dual-branch processing to capture a spectrum of linguistic features. The local branch employs depthwise and pointwise convolutions to distill nuanced, high-frequency local features, while the global branch leverages a downsampled self-attention mechanism to extract infrequent yet critical global context. This bidirectional feature extraction is further refined through a gating mechanism that selectively enhances salient information, resulting in a comprehensive representation that bridges the granularity of local expressions with the broader discourse context.
- **InteractiveNet Fusion Module: Enhanced feature recognition.** Our InteractiveNet Fusion Module is a pivotal enhancement that enriches dialogue comprehension by synergistically integrating CloBlock, a gating mechanism, and Biaffine attention to consolidate multifaceted dialogue features. By carefully orchestrating the interplay between thread, speaker, and reply-specific representations, this module adeptly filters out spurious data noise, thereby distilling a purified and contextually rich set of features. This fusion method not only enhances the signal-to-noise ratio but also amplifies the interpretability of the extracted sentiment elements.
- **Empirical Validation.** Through rigorous empirical evaluation on a suite of benchmark datasets, our proposed IFusionQuad model demonstrates superior performance across a myriad of metrics, cementing its position as the state-of-the-art in dialogue-based sentiment analysis. Specifically, the model achieved a remarkable increase of 6.59% in micro F1 score and a 7.05% improvement in identification F1 for Chinese sentiment quadruple extraction. In the English dataset, the enhancements were equally impressive, with a 2.65% increase in micro F1 and a 4.69% increase in identification F1 scores. The model's proficiency in accurately extracting sentiment quadruples, coupled with its robustness against

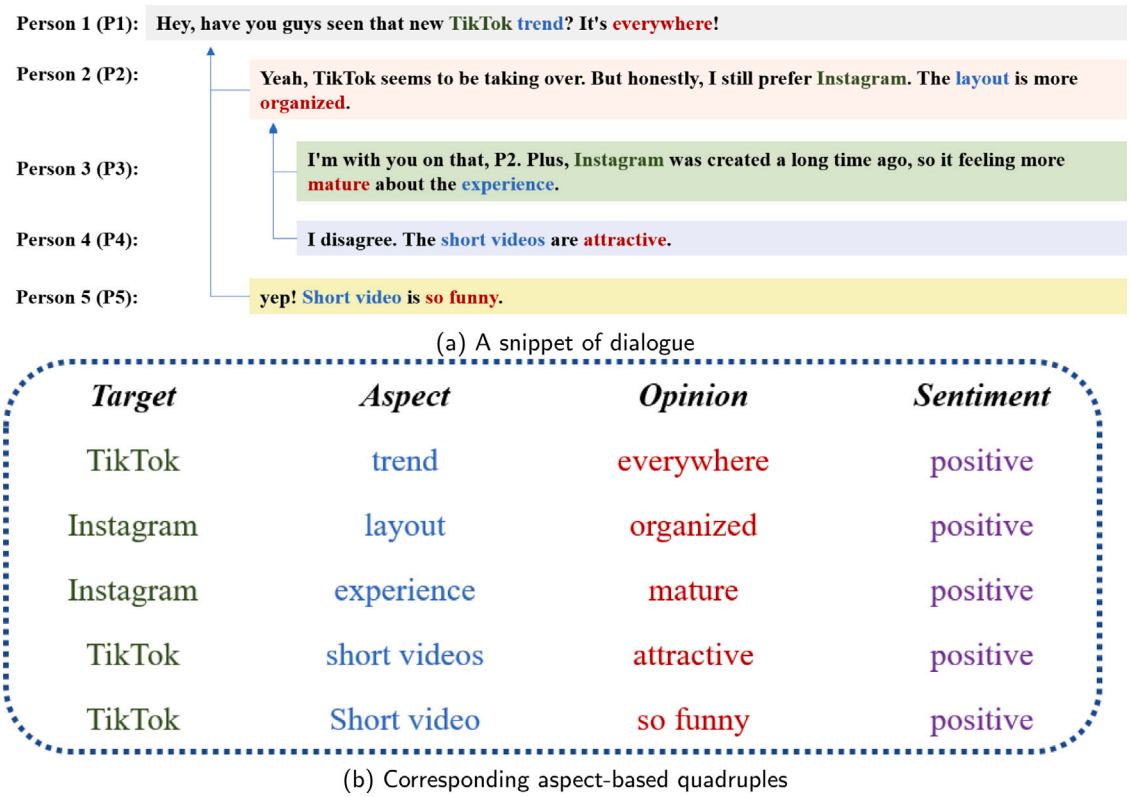


Fig. 1. Visualization depicting the conversational aspect-based sentiment quadruple extraction (DiaASQ). Targets are highlighted in green, aspect terms in blue, opinion terms in red, and sentiment polarity in purple.

various linguistic nuances and conversational complexities, underscores the transformative potential of our approach. This validation not only benchmarks the current efficacy of IFusionQuad but also sets a high watermark for future research endeavours in the domain.

2. Related work

Aspect-based sentiment analysis seeks to detect explicit or implicit sentiments within sentences, a field that has witnessed significant advancements (Consoli et al., 2022; Tang et al., 2019; Xiao et al., 2022). Sentences often contain multiple aspects and opinion terms, reflecting varied sentiment expressions, notably in product and service reviews on e-commerce platforms (Phan & Ogunbona, 2020; Wang et al., 2020a). Extracting opinions from such reviews is vital for merchants to grasp authentic customer feedback, leading to substantial research endeavours in this area.

ABSA methodologies have undergone a transformative evolution, progressing from lexicon-based (Zhang et al., 2019) and machine-learning approaches (Kiritchenko et al., 2014) to the sophisticated realm of deep learning (Dai et al., 2021). Lexicon-based methods, which rely on predefined sentiment lexicons, were among the earliest approaches. These methods, while straightforward, required extensive manual effort to curate comprehensive lexicons. Machine learning approaches, such as those utilizing bag-of-words (Wang et al., 2014) representations and part-of-speech tagging (Gui et al., 2017), marked a significant improvement by automating feature extraction. These strategies, although labor-intensive, demonstrated considerable effectiveness in capturing sentiment nuances.

The advent of deep learning has revolutionized ABSA, offering a transformative shift in contextual representation and task performance. Deep learning techniques, particularly that leverage pre-trained language models like BERT, have introduced superior contextual modelling capabilities. This advancement has led to enhanced performance and efficiency in task-specific applications (Tang et al., 2020; Xiao et al., 2021). For instance, models such as BERT have been instrumental in capturing intricate dependencies within sentences, thus enhancing the accuracy of sentiment detection (Chen et al., 2022; Li et al., 2019b; Sun et al., 2019).

2.1. Single-text aspect-based sentiment analysis

2.1.1. Definition of each task

Currently prevalent ABSA tasks and their variations generally focus on several components or combinations, including aspect terms, sentiment polarity, opinion terms, and aspect categories. Fig. 2, provides a straightforward example to elucidate various traditional single-text ABSA tasks. The figure offers an example to illustrate the subtle differences among various ABSA subtasks. For instance, in the sentence "It looks good, but I decided against buying it because my friend informed me that the battery life of iPhone 12 is poor". We identified two aspect terms, "looks" (a_1) and "battery life" (a_2), associated with the aspect category "iPhone 12" (c). Corresponding opinion terms include "good" (o_1) and "poor" (o_2). The sentiment polarity for these aspects is labelled as positive and negative for (s_1 and s_2) respectively.

The initial ABSA tasks primarily focused on single-output subtasks, aiming to extract specific elements such as aspect (a), sentiment (s), or opinion (o). They involved classifying the sentiment polarity for a given aspect. These subtasks, encompassing Aspect Extraction (AE), Opinion Extraction (OE), Aspect-Level Sentiment Classification (ALSC), and Aspect-Opinion Extraction (AOE), each contribute distinctively to the comprehensive understanding of sentiment analysis. Fig. 2 adeptly illustrates the four distinct subtasks. AE and OE involve extracting aspect and opinion terms (a_1, a_2 and o_1, o_2) from the input sentence, respectively. ALSC determines sentiment towards each aspect by considering the aspect term and sentence as input ($s+a_1 \rightarrow s_1$, $s+a_2 \rightarrow s_2$).

AOE focuses on identifying the opinion term that relates to a specific aspect ($s + a_1 \rightarrow o_1$, $s + a_2 \rightarrow o_2$).

As single-output subtasks matured, the ABSA task shifted focus to compound output subtasks, such as Aspect Extraction and Sentiment Classification (AESC), Pair Extraction (Pair), Aspect Sentiment Triplet Extraction (ASTE), and Aspect Sentiment Quadruple Extraction (ASQE). The four compound subtasks are illustrated in Fig. 2. AESC jointly extracts aspects and classifies sentiments ($(a_1 + s_1)$, $(a_2 + s_2)$). Pair Extraction (*Pair*) identifies the (aspect, opinion) pairs present ($(a_1 + o_1)$, $(a_2 + o_2)$), ASTE captures the (aspect, opinion, sentiment) triplets ((a_1, o_1, s_1), (a_2, o_2, s_2)). Finally, ASQE captures the complete (category, aspect, opinion, sentiment) quadruple. This exploration represents a substantial advancement in comprehensively understanding the emotional nuances in textual data.

2.1.2. Compound output subtask

In the tasks mentioned above, ALSC, AESC, ASTE, and ASQE sentiment polarity classifications are essential for the given aspects. Such scenarios dominate a significant portion of the ABSA task. In Tang et al. (2016) two target-dependent long short-term memory (LSTM) models were created to classify sentiment polarities for given aspects. This is an innovative method that employs LSTM networks to capture the relationships between aspects and their associated sentiments, thereby enhancing the accuracy of sentiment classification. Fan et al. (2023) proposed a Multi-Granularity Attention Network (MGAN) model for aspect sentiment classification. This model employs an attention mechanism to identify word-grained level interactions between aspects and their context. The MGAN model marks a notable progression by offering a deeper understanding of the relationship between aspects and the surrounding context. In Li et al. (2019a), an end-to-end solution was introduced for target-based full sentiment analysis tasks. They proposed a novel unified model employing a unified tagging scheme, utilizing two stacked recursive neural networks. This approach provides a comprehensive framework for sentiment analysis by integrating aspect extraction and sentiment classification into a single model. These methods offer new perspectives and insights into sentiment polarity classification for specific aspects, offering valuable assistance for further research.

Building on the robust research foundation in sentiment polarity classification for specific aspects, compound ABSA-related tasks, such as Pair Extraction and Aspect Sentiment Triplet Extraction (ASTE), have emerged as significant areas of study. These tasks aim to provide a more comprehensive understanding of sentiment by taking into account multiple elements simultaneously. In Zhao et al. (2020), a multi-task learning framework was introduced for Pair Extraction, utilizing shared spans to extract terms under span boundary supervision in an end-to-end approach. This innovative method leverages the shared spans to enhance extraction accuracy, demonstrating the effectiveness of multi-task learning in handling complex ABSA tasks. In Peng et al. (2020) the ASTE task was formally introduced, proposing a two-stage pipeline framework that encompasses aspect extraction, aspect sentiment classification, and opinion extraction. This comprehensive approach laid the groundwork for subsequent research by providing a structured methodology for tackling the ASTE task. In Wu et al. (2020), a grid tagging scheme was presented to facilitate the end-to-end extraction of aspect sentiment triplets, conceptualizing the ASTE task as a grid-based problem. This novel tagging scheme simplifies the extraction process and enhances the model's ability to capture complex sentiment patterns. In Mao et al. (2021), ASTE was recast as a machine reading comprehension task, employing a common BERT encoder to extract triplets across several decoding phases. This approach demonstrates the adaptability of machine reading comprehension techniques in extracting sentiment-related information. In Wu et al. (2021), the authors introduced a syntax fusion encoder that utilizes a Label-aware Graph Convolutional Network (LAGCN) to capture syntactic features and a local attention mechanism to encode parts-of-speech tags, as well as to identify term boundaries. For the aspect-opinion term pairing phase, they implemented Biaffine and Triaffine scoring methods. Additionally, they utilized LAGCN's syntactically enriched representations to enhance scoring with syntax awareness. This approach not only extracts pairs effectively but also generates predictions that are easier to interpret, underscoring the value of incorporating syntactic insights into ABSA models.

The aforementioned approaches have significantly advanced ASTE tasks by introducing methodological innovations. These innovations include the significance of end-to-end approaches, the effective exploitation of word relationships within sentences, the introduction of novel tagging schemes, enhancements to attention mechanisms, and other pioneering developments. These studies provide a foundation for creating more advanced and comprehensive models for sentiment analysis.

2.1.3. Innovative methods

Recent advancements in ABSA research have introduced innovative methods that further extend the capabilities of sentiment analysis models. For example, meta-based self-training with a meta-weighter (MSM) has shown promise in addressing issues such as label imbalance and task-specific learning difficulties. By leveraging a meta-weightier to balance class labels and regulate learning, this approach achieves state-of-the-art results using less labelled data, thus demonstrating the efficiency and adaptability of meta-learning in ABSA (He et al., 2023). Moreover, systematic studies on the biases of pre-trained language models (PLMs) have provided crucial insights into prompt-based sentiment analysis and emotion detection. These studies reveal how PLMs exhibit biases related to label classes, emotional label-word selection, and prompt templates, highlighting areas for improvement in the use of PLMs for affective computing (Mao et al., 2023). The integration of multimodal data has also seen significant advancements. The Atlantis framework, for instance, introduces an aesthetic-oriented multiple granularity fusion network to bridge the representational gap between textual and visual modalities, emphasizing the aesthetic attributes of images. This approach not only enhances the model's ability to identify aspect-sentiment pairs but also sets a new state-of-the-art performance in multimodal ABSA tasks (Xiao et al., 2024). In addition, DiffusionABSA utilizes a novel diffusion model to incrementally refine aspect boundary detection through a denoising process, which effectively handles the colloquial expressions often found in user-generated content. This technique significantly improves the precision of aspect boundary identification, demonstrating the value of diffusion models in sentiment analysis (Liu et al., 2024). Lastly, CoolNet's cross-modal fine-grained alignment and fusion network showcases how advanced alignment between visual clues and textual representations can enhance sentiment polarity prediction. By dynamically integrating semantic and syntactic information, CoolNet achieves superior performance on multimodal sentiment analysis benchmarks, underlining the critical role of cross-modal fusion in improving sentiment analysis accuracy (Xiao et al., 2023). These innovative approaches collectively create opportunities for exploring more sophisticated and comprehensive models, such as the IFusionQuad framework, which aims to integrate dialogue-specific features and enhance sentiment comprehension in conversational contexts.

As sentiment analysis models become increasingly sophisticated, researchers are turning their attention to more complex ABSA tasks. These tasks aim to capture more nuanced and comprehensive sentiment information, thereby enhancing the overall understanding of textual data. In Cai et al. (2021), two novel sentiments quadruplet datasets were annotated, representing a significant step forward in ASQE research. Their provision of these datasets establishes a vital foundation for future studies within the ASQE field. Furthermore, they developed a set of pipeline baselines by incorporating existing models, creating a structured framework for subsequent research to expand upon. Building on this foundation, Zhang et al. (2021) proposed an

innovative paraphrase modelling approach that transforms the ASQE task into a process of paraphrase generation. This innovative approach facilitates end-to-end resolution of ASQE, reducing potential error propagation in pipeline solutions. By training to produce sentiment elements in natural language, this method improves the model's capacity to capture complex sentiment connections. Liu et al. (2021) proposed the Comparative Opinion Quintuple Extraction (COQE) task, which aims to identify comparative sentences and extract all comparative opinion quintuples (Subject, Object, Comparative Aspect, Comparative Opinion, Comparative Preference) from product reviews. They curated three datasets specifically for this task, providing valuable resources for future research. This work extends the scope of sentiment analysis by focusing on comparative opinions, thereby offering a more nuanced understanding of user preferences and sentiments.

The quadruples from the ASQE task offer more comprehensive opinion details than those of previous tasks, thereby better facilitating downstream task development. Building on the ASQE framework, the DiaASQ task substitutes category elements with target elements and reconstructs the dataset to reflect real-world conversational contexts. More importantly, DiaASQ focuses on conversations rather than individual sentence fragments, thereby capturing the dynamic and interactive nature of real-world dialogues. This shift in focus marks a major step forward in sentiment analysis, enabling a more realistic and thorough comprehension of user sentiments in conversational exchanges.

2.2. Recognizing emotions in conversations

Emotion recognition in conversation (ERC) is a subtask of sentiment analysis that aims to identify the emotions expressed in individual utterances within conversations (Fu et al., 2023; Jbene et al., 2022; Zhang et al., 2023). Emotions, recognized as personal psychological states intertwined with thoughts, feelings, and behaviours, are frequently expressed in natural language. This characteristic has bolstered the prominence of ERC in the NLP domain. The ERC facilitates the extraction of sentiments from extensive conversational datasets sourced from platforms such as Facebook, YouTube, Reddit, and Twitter. Its potential extends across diverse domains, including healthcare (for psychological assessment), education (to gauge student frustration), and beyond. ERC research highlights the importance of conversation context modelling, speaker-dependency, and feature integration in multimodal data environments. conversational context modelling often involves the use of Graph Neural Networks (GNN) (Hochreiter & Schmidhuber, 1997), Long Short-Term Memory (LSTM) (Scarselli et al., 2009), and other architectures to improve the understanding of speaker relationships. Majumder et al. (2018) initially used Recurrent Neural Networks (RNNs) for ERC, demonstrating the importance of capturing sequential dependencies in conversations. Ghosal et al. (2019) advanced the field by introducing Graph Convolutional Neural Networks (GCNs) for ERC. Their approach leveraged the relational structure of dialogues, capturing the intricate dependencies between utterances. Wang et al. (2020b) introduced a sequence labelling approach utilizing a Conditional Random Field (CRF) layer to capture emotion consistency in dialogues. This approach ensured that the predicted emotions were coherent across the conversation, addressing the challenge of maintaining emotional consistency in ERC. Zhang et al. (2023) introduced Dual Graph Attention Networks (DualGATs), which consider both discourse structure and speaker-aware context simultaneously. This innovative approach yielded more precise ERC results by integrating multiple aspects of the conversational context.

Emotion recognition in conversation focuses on classifying sentiment from a specified set, whereas tasks such as DiaASQ aim to extract detailed sentiment elements from conversations. Both tasks require a thorough exploration of context-related dialogue features. Related studies provide valuable insights into the DiaASQ task, highlighting the importance of conversational context modelling, speaker dependency, and feature integration in understanding conversational emotions.

3. Method

This section outlines the methodology underlying the proposed framework, designed to address the challenges of DiaASQ task. The exposition commences with a precise formulation of the problem, setting the stage for a deeper exploration of each component of the model.

Firstly, '*Problem Formulation*' provides the theoretical foundation and context necessary for the approach. This is followed by the introduction of the '*Enhanced Labelling Scheme for Grid-Tagging Task Modelling*', which elaborates on the innovative labelling mechanism employed to effectively model the grid-tagging task. This enhanced scheme is pivotal in accurately capturing the nuances of the data, thereby improving model performance. The subsequent section details the '*Input and Encoding Layer*'. This elucidates how PLMs are employed to generate foundational semantic features. This stage serves to prepare the input data for subsequent processing, thereby ensuring that the model is endowed with a understanding and nuanced comprehension of the text. The '*InteractiveNet Fusion*' subsection provides an in-depth explanation of the novel method for combining these features through a sophisticated fusion network.

We then examine the '*Integrating Dialogue Relative Distance*' component, which incorporates contextual information relative to the dialogue, thereby further refining the model's understanding of the interactions within the text. After this, the '*Quadruple Decoding*' section introduces our approach to decoding the quadruple relationships, which is a critical aspect of accurately tagging the grid. To guarantee robust performance, we discuss the '*Loss Function*' designed to optimize the training process, followed by a detailed explanation of our '*Model Training*' procedure. This includes a step-by-step guide on how the model is trained to achieve optimal performance. Finally, we present an '*Algorithm Complexity Analysis of Training*', which includes an in-depth examination of the computational complexity linked to our training algorithm.

The overall architecture and interconnections of these components are depicted in Fig. 3. Each subsection is crafted to build upon the previous one, providing a coherent and thorough understanding of our methodology.

3.1. Problem formulation

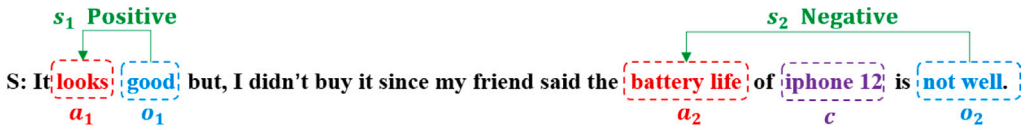
The DiaASQ task focuses on decomposing a given dialogue into its constituent quadruples $Q = \{t, a, o, s\}$, where t, a, o, and s indicate "Target", "Aspect", "Opinion", and "Sentiment Polarity", respectively. Specifically, let us consider a dialogue $D = \{u_i, \ldots, u_n\}$, where u_i represents the ith utterance in the dialogue. Alongside this, there is a corresponding replying record $R = \{r_i, \ldots, r_n\}$, where

- $r_i = 0$ signifies that the ith utterance $u_i \in D$ does not respond to any other utterance within the dialogue.
- $r_i = j$ $(1 \leq j < i)$ signifies that the ith utterance $u_i \in D$ is respond to the jth utterance $u_j \in D$ within the dialogue.

The objective of the DiaASQ task is to determine the quadruples that make up the dialogue. Let the notation $\{w_{i1}, \ldots, w_{im}\}$ represent the words in u_i, where m represents the overall count of words in that utterance and each element $\{u_i\} = \{w_{i1}, \ldots, w_{im}\}$ in the set D is a text segment. The set D itself comprises all such utterances within the dialogue. The replying record R captures the hierarchical tree structure of the dialogue D. This structure is essential for understanding the relationships and interactions between the different utterances in the dialogue.

For a more detailed explanation, it should be noted that the DiaASQ task comprises the following components:

(1) A dialogue $D = \{u_i, \ldots, u_n\}$;
(2) A reply record $R = \{r_i, \ldots, r_n\}$;



Subtasks	Input	Output	Task Type
Aspect Term Extraction (AE)	S	a_1, a_2	Extraction
Opinion Term Extraction (OE)	S	o_1, o_2	Extraction
Aspect-level Sentiment Classification (ALSC)	S + a_1	s_1	Classification
	S + a_2	s_2	
Aspect-oriented Opinion Extraction(AOE)	S + a_1	o_1	Extraction
	S + a_2	o_2	
Aspect Term Extraction and Sentiment Classification(AESC)	S	(a_1 + s_1)	Extraction &Classification
		(a_2 + s_2)	
Pair Extraction(Pair)	S	(a_1 + o_1)	Extraction
		(a_2 + o_2)	
Aspect Sentiment Triplet Extraction(ASTE)	S	(a_1, o_1, s_1)	Extraction &Classification
		(a_2,o_2, s_2)	
Aspect Sentiment Quadruple Extraction(ASQE)	S	(c,a_1, o_1, s_1)	Extraction &Classification
		(c,a_2,o_2, s_2)	

Fig. 2. Illustration of Eight Single-Text Aspect-Based Sentiment Analysis Subtasks.

(3) Identification of quadruples that represent the structural and semantic relationships within the dialogue;
(4) Each utterance u_i consists of a set of words $\{w_{i1}, \ldots, w_{im}\}$, with m being the length of the utterance.

To provide further insight into the complexities of the task, a comprehensive example is offered. Consider the following dialogue and its associated response record:

- **Dialogue(D)**:
  < **1** >: "Hello, the weather is nice today". (Length: 7)
  < **2** >: "Yes, it's sunny and I'm in a good mood too". (Length: 12)
  < **3** >: "What are you planning to do?" (Length: 8)
  < **4** >: "I plan to go for a walk and enjoy the sunshine". (Length: 11)
- **Replying Record(R)**:
  < **1** >: $r_1 = 0$ means that the first sentence is not a reply to any other utterance;
  < **2** >: $r_2 = 1$ means that the second sentence is a reply to the first sentence;
  < **3** >: $r_3 = 0$ means that the third sentence is not a reply to any other utterance;
  < **4** >: $r_4 = 3$ means that the fourth sentence is a reply to the third sentence.

Given the input D and R, DiaASQ generates all sentiment-oriented quadruples $Q = \{t, a, o, s\}$. The sentiment polarity may fall into one of three categories: {positive, negative, other}. For instance, as illustrated in Fig. 1, one of the extracted quadruples $(TikTok, trend, everywhere, positive)$ demonstrates the model's capability to discern aspect targets and diverse sentiment orientations within a dialogue.

3.2. Enhanced labelling scheme for grid-tagging task modelling

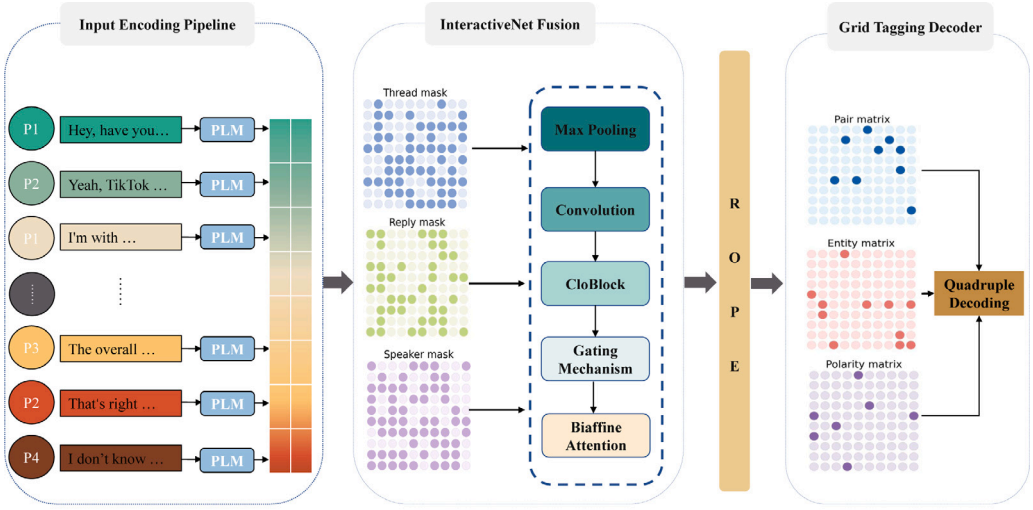
The DiaASQ task requires simultaneous extraction of the target, aspect terms, and opinion terms. Building on prior grid annotation methods (Cai et al., 2021; Zhang et al., 2021), the grid labelling scheme has been refined to meet the specific requirements of this task. As illustrated in Fig. 4, the task is divided into three distinct collaborative endeavours: entity boundary detection, entity pair detection, and sentiment polarity detection.

- **Entity Boundary Labels**: The labels are employed to denote the starting and ending points of token-level relations for target, aspect, and opinion terms. The labels 'tgt', 'asp', and 'opi' are used for this purpose. Fig. 4 illustrates the labelling scheme. The term 'tgt' signifies the target term 'iPhone 14' between 'iPhone' and '14', whereas 'asp' indicates the aspect term 'battery quality' between 'battery' and 'quality'. Similarly, 'opi' denotes the opinion term 'very good'. This labelling scheme ensures precise identification of the boundaries of each entity type.
- **Entity Pair Labels**: Once entity boundaries have been delineated, various term types are merged into composites using entity pair labels. The labels 'h2h' and 't2t' are utilized to synchronize head and tail markers for entity pairs of different types. As illustrated in Fig. 4, the term 'iPhone' (target) is linked to the term 'battery' (aspect) via the label 'h2h', while the term '14' (target) is connected to the term 'quality' (aspect) via the label 't2t'. The process of tagging term pairs of distinct types generates triplets (t_k, a_k, o_k) to represent entity relationships. This step is crucial for the accurate representation of the interactions between different entities within the dialogue.
- **Sentiment Polarity Labels**: Following the acquisition of a triplet, it is expanded to a quadruplet, $q_k = (t_k, a_k, o_k, p_k)$, by appending an emotion category label, 'p_k'. The emotion polarity is uniquely influenced by both the target and the opinion, therefore, the labels are positioned between the beginnings and endings of these terms. Fig. 4 illustrates the manner in which the emotion label 'pos' is linked to the target and opinion terms. In this example, the target term 'iPhone' is linked with the opinion term 'very' at the head, while the target term '14' is linked with the opinion term 'nice' at the tail. This labelling mechanism ensures that the sentiment polarity is accurately captured, reflecting the nuanced emotional context of the dialogue.

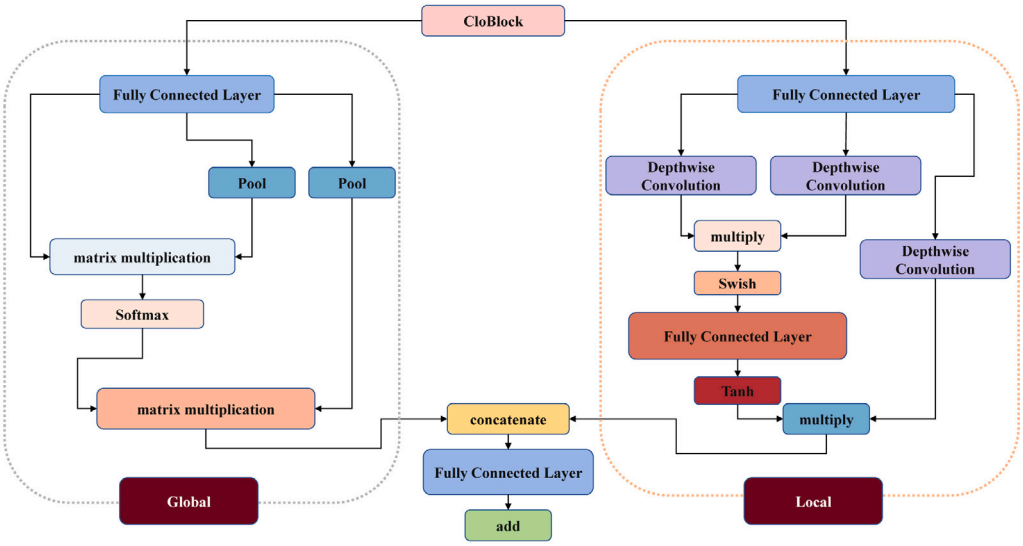
The key innovative technologies that contribute to this enhanced labelling scheme include the precise delineation of entity boundaries, effective merging of entity pairs, and the accurate labelling of sentiment polarity.

3.3. Input and encoding layer

The input and encoding layers are the pivotal components of the proposed framework. Its function is to transform raw dialogue data into



(a) The overall framework of our IFusionQuad model.



(b) The CloBlock component of IFusionQuad model.

Fig. 3. Architectural Diagram of the IFusionQuad Model, with Component Cloblock Detailed in (b) within (a).

a format suitable for subsequent processing. This layer leverages PLMs, such as BERT, to generate foundational semantic features.

Initially, each utterance $u_i \in D$ is tokenized into individual words or subwords. To address the potential issue of dialogue length exceeding the model's maximum input length, each utterance is encoded separately using its own PLM. The special tokens $[CLS]$ and $[SEP]$ are employed to delineate each utterance, as illustrated in Eq. (1):

$$u_i' = \langle [CLS], w_1, \ldots, w_m, [SEP] \rangle, \tag{1}$$

where, u_i' represents the tokenized sequence containing the special tokens $[CLS]$ and $[SEP]$.

The tokenized sequence u_i' is then fed into the PLM to obtain contextual embeddings. The PLM processes the sequence and produces a set of embeddings H_i' for each token, including the special tokens $[CLS]$ and $[SEP]$, as depicted in Eq. (2):

$$H_i' = h_{\text{cls}}, h_1, \ldots, h_m, h_{\text{sep}} = \text{PLM}(u_i'), \tag{2}$$

In this equation, the symbol H_i' denotes the contextual embeddings obtained from the PLM applied to the utterance token u_i'. These embeddings encapsulate the surrounding information for each token within the utterance, thereby enabling the model to comprehend the meaning of each token in relation to its context.

3.4. InteractiveNet fusion

The InteractiveNet fusion module was designed to improve the capacity of the model to comprehend utterances in dialogue. The architecture consists of two primary components, namely the Interactive Dialogue Component and the Feature Fusion Module. The objective of the Interactive Dialogue Component is to identify features that are unique to each utterance, while the Feature Fusion Module is responsible for integrating these features into a coherent whole.

Interactive Dialogue Component: The component employs dialogue threads, speakers, and replies as the fundamental features of a conversation. The creation of an attention mask, M^c, and subsequent integration of biases from preceding features allows for control over the interaction of tokens within the dialogue. Here, $c \in \{Th, Sp, Rp\}$ represents the various types of token interactions, namely topic (Th),

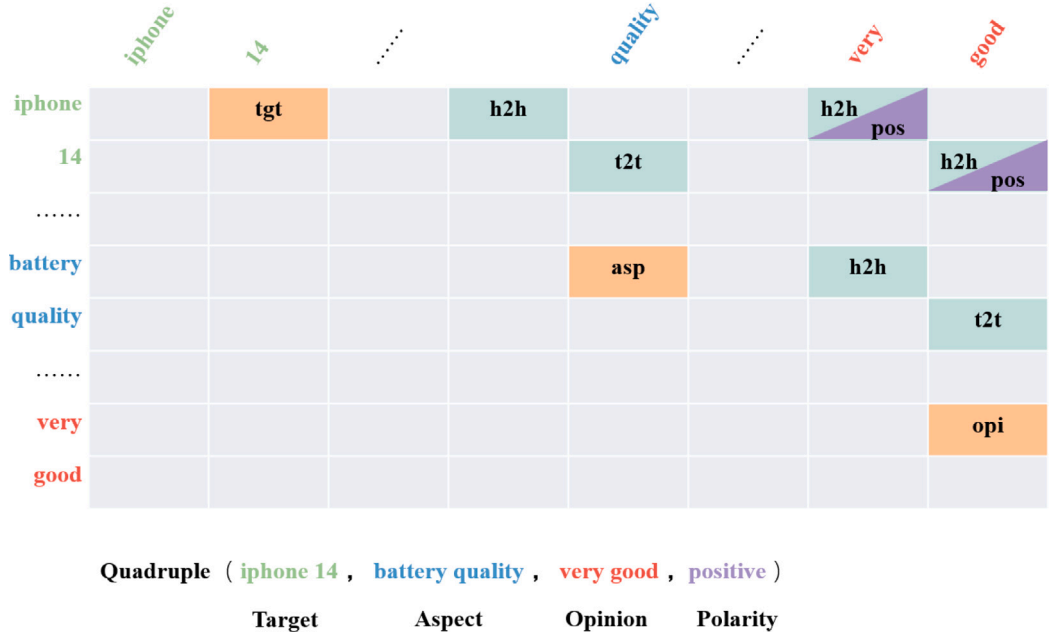


Fig. 4. Refined labelling scheme for DiaASQ task, depicting the collaborative efforts in entity boundary detection, entity pair detection, and sentiment polarity detection.

speaker (Sp), and reply (Rp). The principal objective of this methodology is to introduce masks for different interaction types, thus enabling precise control over token interactions.

$$H^c = \text{Masked-Att}(Q, K, V, M^c) = \text{SoftMax}\left(\frac{(Q^T \cdot K) \odot M^c}{\sqrt{D}} \cdot V\right), \quad (3)$$

Here, $Q = K = V = H \in \mathbb{R}^{N \times d}$ represents the complete sequence formed by concatenating token embeddings for each utterance (denoted as H_i in Eq. (2)), where N is the length of token for D while $\odot$ denotes element-wise multiplication. The matrix $M^c \in \mathbb{R}^{N \times N}$ is defined as follows:

- **Thread Mask**: $M_{ij}^{Th} = 1$ when the ith and jth tokens are part of the same dialogue thread.
- **Speaker Mask**: $M_{ij}^{Sp} = 1$ when the ith and jth tokens originate from the same speaker.
- **Reply Mask**: $M_{ij}^{Rp} = 1$ when the utterances containing the ith and jth tokens, respectively, are in a replying relationship.

Then max-pooling is performed on the mask representations:

$$H_f = \text{Max-Pooling}(H^{\text{Th}}, H^{\text{Sp}}, H^{\text{Rp}}), \quad (4)$$

The **Feature Fusion Module** was designed to integrate various types of features to more comprehensively capture the characteristics of specific utterances. Initially, the mask is input into our meticulously designed CloBlock. Drawing inspiration from Fan et al. (2023), as illustrated in Fig. 3(b), each CloBlock comprises a local branch and a global branch.

- **Global Branch**: The global branch employs a conventional attention mechanism that incorporates the downsampling of K and V to mitigate computational complexity. This approach facilitates the capture of low-frequency global information. The specific formulation is as follows:

  $$X_{\text{global}} = \text{Attention}(Q_g, \text{Pool}(K_g), \text{Pool}(V_g)), \quad (5)$$

  Although it is proficient at capturing low-frequency global information, the model lacks the capability to process high-frequency local information. To overcome this constraint, the integration of the Local Branch is essential for enhancing the model's performance.

- **Local Branch**: In Eq. (6) and Eq. (7), X_{in} represents the input to the Local Branch, with FC denoting a fully connected layer. After this linear transformation, we initiate a process of local feature aggregation on V using shared weights. Subsequently, we leverage the processed V along with Q and K to perform a context-aware local enhancement.

  $$X_{\text{in}} = \textbf{Convolution}(H_f) \quad (6)$$

  $$Q, K, V = \text{FC}(X_{\text{in}}), \quad (7)$$

  In Eq. (8), a straightforward depthwise convolution (DWconv) is employed to collect local information for V with the DWconv weights being shared globally.

  $$V_s = \text{DWconv}(V), \quad (8)$$

  After integrating the locally aggregated information on V with the shared weights, Q and K are combined to generate context-aware weights. It is noteworthy that we adopt an approach that differs from local self-attention, which tends to be more intricate. Specifically, we initially aggregate the local information of Q and K separately using two DWconvs. Subsequently, we compute the Hadamard product of Q and K, followed by a series of transformations to obtain context-sensitive weights that range from −1 to 1. Finally, the local features undergo enhancement using the derived weights. The overall process is summarized as follows:

  $$Q_l = \textbf{DWconv}(Q), \quad (9)$$

  $$K_l = \textbf{DWconv}(K), \quad (10)$$

  $$Attn_t = \textbf{FC}(\textbf{Swish}(\textbf{FC}(Q_l \odot K_l))), \quad (11)$$

  $$\text{Attn} = \textbf{Tanh}\left(\frac{Attn_t}{\sqrt{d}}\right), \quad (12)$$

  $$X_{\text{local}} = \textbf{Attn} \odot V_s, \quad (13)$$

  The outputs of the Global Branch and Local Branch are concatenated along the channel dimension, as indicated in Eq. (14) and

(15). Subsequently, a fully connected layer is applied along the channel dimension.

$$X_t = \textbf{Concat}(X_{\text{local}}, X_{\text{global}}), \tag{14}$$

$$X_{\text{out}} = \textbf{FC}(X_t), \tag{15}$$

Algorithm 1 Procedure of CloBlock

Input: Input feature map H_f
Output: Output feature map X_{out}
function CloBlock(H_f)
  Global Branch:
  $Q_g, K_g, V_g \leftarrow \text{FC}(H_f)$ ▷ Fully connected layer to generate Q_g, K_g, V_g
  $K_g^{\text{pool}}, V_g^{\text{pool}} \leftarrow \text{Pool}(K_g), \text{Pool}(V_g)$ ▷ Downsampling K and V
  $X_{\text{global}} \leftarrow \text{Attention}(Q_g, K_g^{\text{pool}}, V_g^{\text{pool}})$ ▷ Global attention calculation
  Local Branch:
  $X_{\text{in}} \leftarrow \text{Convolution}(H_f)$ ▷ Apply convolution to capture local features
  $Q, K, V \leftarrow \text{FC}(X_{\text{in}})$ ▷ Fully connected layer to generate Q, K, V
  $V_s \leftarrow \text{DWconv}(V)$ ▷ Depthwise convolution on V for local information aggregation
  $Q_l \leftarrow \text{DWconv}(Q)$ ▷ Depthwise convolution on Q
  $K_l \leftarrow \text{DWconv}(K)$ ▷ Depthwise convolution on K
  $Attn_t \leftarrow \text{FC}(\text{Swish}(\text{FC}(Q_l \odot K_l)))$ ▷ Hadamard product followed by two fully connected layers
  $\text{Attn} \leftarrow \text{Tanh}\left(\frac{Attn_t}{\sqrt{d}}\right)$ ▷ Apply Tanh normalization
  $X_{\text{local}} \leftarrow \text{Attn} \odot V_s$ ▷ Enhance local features
  Feature Fusion:
  $X_t \leftarrow \text{Concat}(X_{\text{local}}, X_{\text{global}})$ ▷ Concatenate along the channel dimension
  $X_{\text{out}} \leftarrow \text{FC}(X_t)$ ▷ Fully connected layer for final feature map
  return X_{out}
end function

The CloBlock algorithm described in Algorithm 1 is designed to effectively integrate global and local features, enhancing the model's ability to understand both high-level context and fine-grained details within input data. Here's a detailed explanation of the key components and steps involved in the CloBlock algorithm:

- **Global Branch:** The global branch employs a conventional attention mechanism using queries (Q_g), keys (K_g), and values (V_g) generated by a fully connected layer. By incorporating pooling operations on K_g and V_g, the computational complexity is reduced, allowing the model to efficiently capture low-frequency global information. This approach is beneficial because it enables the model to focus on significant global patterns without being overwhelmed by computational demands. The global context is essential for understanding the overall structure and semantics of the input data.
- **Local Branch:** The local branch is designed to complement the global branch by capturing high-frequency local information. It starts with a convolution operation to extract local features from the input feature map (H_f), followed by the use of fully connected layers to produce local Q, K, and V. Depthwise convolution (**DWconv**) is then applied to V to aggregate local information efficiently. The use of shared weights in DWconv helps reduce the number of parameters and computational overhead while retaining the capability to capture localized patterns. Local feature enhancement is further achieved by computing context-aware weights through a series of transformations, including Hadamard product and activation functions. This process ensures that the local features are dynamically adjusted based on the surrounding context, providing a more detailed and accurate representation of local information.
- **Feature Fusion:** After processing through the global and local branches, the respective outputs (X_{global} and X_{local}) are concatenated along the channel dimension. This fusion step combines the strengths of both global and local feature representations, ensuring that the model benefits from both high-level contextual understanding and detailed local information. A fully connected layer is applied to the concatenated features, which helps in integrating the diverse information into a coherent output representation. This step is crucial as it allows the model to leverage the complementary nature of global and local features, thereby improving overall performance in tasks that require both macro and micro-level feature analysis.
- **Non-linear Interaction:** The use of Swish activation functions and Tanh normalization in the computation of attention weights allows the model to capture complex, non-linear interactions between features. These functions help maintain stable gradients during training and ensure that the attention mechanism remains effective across different layers and scales of the network. This design choice contributes to the robustness and reliability of the model, especially when handling intricate patterns and relationships in the input data.

In conclusion, the CloBlock algorithm's dual-branch structure and feature fusion mechanism provide a balanced approach to feature extraction and integration. By effectively combining global and local information, the model achieves a comprehensive understanding of the input data, making it suitable for a wide range of applications that require both contextual and detailed feature analysis.

After processing, a **Gating Mechanism** (refer to Eq. (16)) is applied to selectively filter salient features and alleviate noise interference within the input representation X_{out}. Subsequently, our model integrates a **Biaffine attention** mechanism (Eq. (17)) to produce the feature v_i^c. This selection is based on the demonstrated efficacy of this module in addressing ABSA tasks (Dozat & Manning, 2016).

$$H_{out} = \textbf{GatingMechanism}(X_{\text{out}}), \tag{16}$$

The **Gating Mechanism** applied in Eq. (16) after the CloBlock module serves several critical purposes in enhancing the model's performance. The CloBlock module effectively integrates both global and local features from the input representation, providing a rich and comprehensive feature set. However, not all features extracted are equally relevant or useful for the downstream tasks. The inclusion of a gating mechanism allows the model to selectively filter and prioritize salient features while suppressing less relevant or noisy ones.

The gating mechanism functions as a dynamic filter, enabling the model to control the flow of information based on the context and the specific task requirements. By learning which features to emphasize and which to attenuate, the gating mechanism reduces the risk of overfitting and enhances the robustness of the model against irrelevant noise in the data. This selective feature emphasis is crucial for tasks like Aspect-Based Sentiment Analysis (ABSA), where fine-grained distinctions in language need to be captured accurately.

Moreover, the gating mechanism helps in maintaining a balanced trade-off between retaining valuable information and discarding irrelevant details. This balance improves the quality of the representation passed on to subsequent components, such as the **Biaffine attention** mechanism, which is applied in Eq. (17). The output H_{out} from the gating mechanism provides a refined and focused feature map, which enhances the effectiveness of the Biaffine attention module. This module is particularly adept at capturing complex interactions within the data, further improving the model's ability to handle intricate sentiment relationships.

The gating mechanism after the CloBlock processing ensures that the model can dynamically adapt its feature selection process, leading to more robust, accurate, and interpretable outcomes. This capability is essential for effectively managing the complexity of natural language processing tasks, such as ABSA, where both global context and local details are critical for accurate predictions.

$$v_i^r = \textbf{BiaffineAttention}(H_{out}), \tag{17}$$

in Eq. (17), r ∈ {tgt, … , h2h, … , pos, … , ε_{ent}, …} represents a particular label, ε_{ent} signifies the non-relation label within the entity boundary matrix.

3.5. Integrating relative distance in dialogue

One limitation of Pre-trained language models is that they often require each utterance in a conversation to be encoded separately, which can potentially undermine the coherence of the conversation. To address this limitation, the integration of RoPE (Su et al., 2021) into token representations is proposed. RoPE introduces a novel approach to positional encoding by rotating the embedding space based on the relative positions of tokens. This method enables the model to more accurately capture the relative distances among tokens in contrast to traditional positional encodings. By dynamically encoding the relative distances between utterances, RoPE enhances the model's comprehension and helps maintain the coherence of the conversation by providing crucial distance information. The effectiveness of this approach has been validated in the study by Li et al. (2023). The mathematical formulation of RoPE is given by:

$$u_i^r = \mathcal{R}(\theta, i) v_i^r, \tag{18}$$

Here, $\mathcal{R}(\theta, i)$ is a rotation matrix that encodes the positional information. The parameter θ controls the rotation, and i represents the absolute position (index) of the token v_i^r. This rotation matrix is applied to the token representation v_i^r to obtain the positionally encoded token u_i^r.

The primary benefit of integrating RoPE is the improved ability to capture the relative distances between tokens, which enables the model to gain a deeper understanding of the conversation's context and flow of the conversation, leading to more accurate and coherent representations of dialogue utterances.

Integrating RoPE into the IFusionQuad model offers several key advantages:

1. **Enhanced Contextual Coherence**: By encoding relative positional information directly into token embeddings, RoPE ensures that the model captures not only the content of the tokens but also their contextual relationships within the dialogue. This facilitates the maintenance of coherence and flow in conversations, which is critical for accurately understanding multi-turn dialogues.
2. **Improved Handling of Long Sequences**: Traditional positional encodings often struggle with long sequences, leading to diminished performance. RoPE's approach to rotating embedding spaces allows it to more effectively capture long-range dependencies, making it better suited for tasks involving extended conversations or documents.
3. **Computational Efficiency**: Unlike some other methods that require complex calculations or additional layers, RoPE's mechanism is relatively straightforward, involving only the application of a rotation matrix. This simplicity can lead to computational efficiency, making it suitable for real-time processing scenarios.
4. **Adaptability**: RoPE can be seamlessly integrated with various types of PLMs without substantial architectural modifications. This adaptability makes it a versatile choice for different NLP tasks beyond dialogue modelling.

However, there are also some limitations associated with the integration of RoPE:

1. **Limited by Rotational Symmetry**: While RoPE captures relative positional information effectively, it operates under the assumption of rotational symmetry in positional relationships. This assumption may not always hold, potentially leading to inaccuracies in specific contexts or within nuanced language structures.
2. **Dependency on Predefined** θ: The performance of RoPE is influenced by the choice of the parameter θ. Selecting an optimal θ requires careful tuning and may vary across different datasets and tasks, which can be time-consuming and computationally expensive.
3. **Sensitivity to Long-Range Contexts**: Although RoPE improves the handling of longer sequences, its performance may still degrade in cases where there are extremely long-range dependencies, similar to the limitations observed in other relative positional encoding methods.

To integrate RoPE within the IFusionQuad model, we apply the rotation matrix $\mathcal{R}(\theta, i)$ to the token representations. The implementation involves the following steps:

- **Step 1: Initial Token Embedding**: Each token v_i^r is initially embedded using standard embedding techniques, resulting in a representation that does not yet incorporate positional information.
- **Step 2: Compute Rotation Matrix**: For each token v_i^r at position i, compute the rotation matrix $\mathcal{R}(\theta, i)$, where θ is a predefined parameter that controls the rotation degree based on the token's position.
- **Step 3: Apply Rotation Matrix**: The computed rotation matrix is applied to each token embedding, transforming v_i^r into u_i^r using Eq. (18). This transformation encodes the token content and its relative position within the dialogue.
- **Step 4: Integrate with Attention Mechanism**: The positionally encoded tokens u_i^r are fed into the model's attention mechanism, enhancing the ability to capture relationships and dependencies among different parts of the dialogue based on both content and relative positions.

This implementation enables the IFusionQuad model to effectively incorporate relative distance information between dialogue utterances, resulting in more accurate and context-aware representations. The use of RoPE particularly helps to maintain the sequential integrity of dialogues, which is essential for understanding the nuanced dynamics of conversational data. By leveraging relative positions, the model can better distinguish between contextually significant and less relevant information, thereby improving overall task performance, especially in multi-turn dialogue settings.

3.6. Quadruple decoding method

Based on each tag representation u_i^r, the score for every tag pair is calculated according to the label r.

$$s_{ij}^r = (u_i^r)^T u_j^r, \tag{19}$$

In this context, s_{ij}^r represents the probability of the relationship label r between w_i and w_j. A softmax layer is utilized on every element of each matrix to identify the relationship label r. For instance, the probabilities within the entity boundary matrix are calculated as follows:

$$p_{ij}^{ent} = \text{Softmax}([s_{ij}^{\varepsilon_{ent}}; s_{ij}^{tgt}; s_{ij}^{asp}; s_{ij}^{opi}]), \tag{20}$$

After acquiring all the labels in the grid, we decode the quadruples according to the rules described in Section 3.2.

3.7. Loss function

The goal of training is to reduce the cross-entropy loss associated with each subtask:

$$L_k = -\frac{1}{G \cdot N^2} \sum_{g=1}^{G} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_k y_{ij}^k \log(p_{ij}^k), \tag{21}$$

where $k \in \{\text{ent}, \text{pair}, \text{pol}\}$ denotes a specific subtask, N represents the overall token length within a dialogue, and G indicates the total number of training examples. y_{ij}^k refers to the ground-truth label, and p_{ij}^k signifies the predicted probability. Due to the imbalance in label types (as discussed in Section 3.2), we utilize a specific tag-wise weighting vector α_k to mitigate this issue.

The final loss is the aggregate of the losses for all three subtasks:

$$L = L_{\text{ent}} + \beta L_{\text{pair}} + \eta L_{\text{pol}}, \tag{22}$$

3.8. Model training

In this section, we describe the training process of the IFusionQuad, a sophisticated model designed for the DiaASQ task. The training algorithm is meticulously crafted to optimize the IFusionQuad model, ensuring it accurately extracts Quadruples from textual data. The overall procedure is encapsulated in Algorithm 2 , which we detail below. The training commences with the input of a dialogue set D and its replying record set r of utterance. The objective is to train the IFusionQuad model to output quadruple Q for a given dialogue, thereby enhancing the model's ability to identify quadruples from the dialogue. The core training loop is iterative until the model converges to an optimal performance state. This iterative process is succinctly represented as:

1: **Repeat**
2: train(IFusionQuad (D, r)
3: **Until Convergence**

Upon convergence, the IFusionQuad model is considered trained, and the quadruples Q are extracted using the IFusionQuad model with the dialogue set D and the corresponding replying record set r as inputs. The function **IFusionQuad** encapsulates the core operations of the model. Initially, the function **Encoder** is invoked to generate embeddings H^i from the dialogue set D and the replying record set r. These embeddings are derived by applying a labelling scheme to D, resulting in a modified dialogue set D', and subsequently utilizing a PLM to obtain H^i.

The function **InteractiveNetFusion** processes the embeddings H^i. This involves creating a thread, speaker, and reply mask (M^c), which are then employed in a masked attention mechanism to produce masked attention representations H^c. The resulting representations are combined using max-pooling, integrating the thread, speaker, and reply information into H_f. These combined features are processed through a CloBlock, resulting in X_{out}. The **GatingMechanism** is then applied to X_{out}, producing gated representations H_{out}. These representations are further refined using Biaffine attention, yielding v_i^r. The function **Integrating** then integrates the Biaffine attention representations v_i^r by applying a transformation $\mathcal{R}(\theta, i)$, resulting in u_i^r.

Finally, **QuadrupleDecoding** decodes the integrated representations u_i^r into the desired quadruples Q. This decoding involves calculating a scoring function s_{ij}^r and applying a softmax function to predict the entities, targets, aspects, and opinions, thereby yielding the final quadruples.

The **Encoder**, **InteractiveNetFusion**, **Integrating**, and **QuadrupleDecoding** functions are essential components of the IFusionQuad model, each contributing a unique aspect to its operation. The *Encoder* function utilizes a PLM to generate embeddings for the dialogue set D and replying record set r. The *InteractiveNetFusion* function integrates various contextual features using masked attention, max-pooling, and a gating mechanism. The *Integrating* function transforms the attention representations, while the *QuadrupleDecoding* function computes the final quadruples based on the integrated representations.

This comprehensive training algorithm ensures that the IFusionQuad model is optimally trained to perform the DiaASQ task. By leveraging both semantic and contextual features, the model accurately extracts sentiment-focused quadruples from dialogue data. Through a series of iterative training steps, the algorithm enhances the model's capacity to comprehend and integrate various aspects of dialogue, including speaker roles and reply structures, resulting in precise and reliable extraction of sentiment-oriented information.

Algorithm 2 Overall procedure of IFusionQuad

Input: Dialogue set D and corresponding replying record set r of utterance
Output: Quadruples Q of the given dialogue and Trained model IFusionQuad

repeat
  train(IFusionQuad(D, r)
until Convergence
$Quadruples \leftarrow IFusionQuad(D, r)$,
return Trained model IFusionQuad, Quadruples
function IFusionQuad(D)
  $H^i \leftarrow$ Encoder(D)
  $v_i^c \leftarrow$ InteractiveNetFusion(H_i)
  $u_i^r \leftarrow$ Integrating(v_i^c)
  $Quadruples \leftarrow$ QuadrupleDecoding(u_i^r)
  return $Quadruples$,Model IFusionQuad
end function
function Encoder(D, r)
  $D' \leftarrow$ labeling scheme (D)
  $H^i \leftarrow$ PLM(D',)
  return H^i
end function
function InteractiveNetFusion(H^i)
  $M^c \leftarrow$ Thread,Speaker and Reply Mask (Section 3.4)
  $H^c \leftarrow$ Masked-Att(Q, K, V, M^c)
  $H_f \leftarrow$ Max-Pooling($H^{\text{Th}}, H^{\text{Sp}}, H^{\text{Rp}}$) (Equation (4))
  $X_{out} \leftarrow$ CloBlock(Q, K, V, H_f)(Equation (5) - Equation (15))
  $H_{out} \leftarrow$ Gating Mechanism(X_{out}) (Equation (16))
  $v_i^r \leftarrow$ BiaffineAttention(H_{out})(Equation (17))
end function
function Integrating(v_i^r)
  $u_i^r \leftarrow \mathcal{R}(\theta, i) v_i^r$ (Equation (18))
end function
function QuadrupleDecoding(v_i^r)
  $s_{ij}^r \leftarrow (u_i^r)^T u_j^r$ (Equation (19))
  $p_{ij}^{ent} \leftarrow$ Softmax($[s_{ij}^{\varepsilon_{ent}}; s_{ij}^{tgt}; s_{ij}^{asp}; s_{ij}^{opi}]$) (Equation (20))
end function

3.9. Algorithm complexity analysis of training

The time complexity analysis of the IFusionQuad involves a detailed examination of each component and operation within the Algorithm 2. This analysis aids in understanding the computational efficiency and scalability of the proposed method.

- **Encoding Layer**: The encoding layer utilizes a PLM to transform the input dialogue set D into hidden representations H^i. The time complexity of the PLM, such as BERTis influenced by the token count in the input dialogues as well as the model's architecture,(specifically the number of layers L, hidden units H, and self-attention heads A). This results in a complexity of $O(nLHA)$, where n is the mean length of the dialogues within the dataset D.

- **InteractiveNetFusion**: The InteractiveNetFusion component processes the embeddings H^i. This involves creating thread, speaker, and reply masks (M^c), applying masked attention, performing max-pooling, and processing through a CloBlock. The masked attention mechanism has a complexity of $O(n^2A)$ due to interactions between tokens. Max-pooling and subsequent operations like CloBlock and gating mechanism add a complexity proportional to the input size, approximating to $O(n)$. Therefore, the total complexity for InteractiveNetFusion is $O(n^2A)$.
- **Integrating**: The Integrating function applies a transformation $\mathcal{R}(\theta, i)$ to the Biaffine attention representations v_i^r. A linear transformation over each token representation is $O(nd)$., where d is the dimensionality of the representations.
- **QuadrupleDecoding**:The QuadrupleDecoding function decodes the integrated representations u_i^r into quadruples. This involves calculating a scoring function s_{ij}^r for each pair of tokens, leading to a complexity of $O(n^2)$. The softmax function applied subsequently maintains this complexity.

Summing up all components, the time complexity of the IFusionQuad training algorithm can be approximated as follows:

- Encoding Layer: $O(nLHA)$
- InteractiveNetFusion: $O(n^2A)$
- Integrating: $O(nd)$
- QuadrupleDecoding: $O(n^2)$

Since $O(n^2A)$ and $O(n^2)$ terms are likely to dominate for longer dialogues and larger feature dimensions, the time complexity can be further approximated as $O(nLHA)+O(n^2A)+O(nd)+O(n^2)$. Simplifying, we get the dominant terms as $O(nLHA)+O(n^2A)$. Thus, the total time complexity can be approximated as:

$$O(nLHA)+O(n^2A)+O(nd)+O(n^2)\approx O(nLHA)+O(n^2A)$$

It is important to note that while the time complexity per dialogue is quadratic with respect to the dialogue length, in practice, the lengths of dialogues are typically bounded. Additionally, implementation efficiencies—particularly those related to matrix operations in PLMs and attention mechanisms—can significantly mitigate computational costs. Nonetheless, the scalability of IFusionQuad may still be challenged by very large datasets or exceedingly long dialogues, necessitating efficient batching, parallelization, and hardware acceleration techniques for practical deployment.

4. Experiment

This section presents the results of the experiments conducted, which are displayed through tables and figures, accompanied by comprehensive analyses. Initially, we present the commonly used DiaASQ datasets along with the relevant experimental configurations. Following this, a thorough presentation of the experimental results is provided.

4.1. Datasets

To effectively tackle the DiaASQ task, a robust dataset is essential. Li et al. (2023) constructed a novel dataset using data from Weibo, China's largest social media platform, which is known for its diverse and dynamic conversational content. This dataset comprises 9 million posts and comments from 100 verified digital bloggers, forming multi-threaded, multi-round conversations with a tree structure (Fig. 5). After extensive preprocessing and rigorous cleaning, a final selection of 1000 conversations was made, ensuring a high-quality dataset that accurately reflects real-world dialogue dynamics. These dialogues were annotated by crowd-sourced workers trained under SemEval ABSA guidelines (Pontiki et al., 2014). To guarantee annotation quality, linguistic and computer science experts reviewed the annotations. The annotations were cross-checked and verified using automatic rules, achieving a high inter-annotator agreement with a Cohen's Kappa score of 0.86.

As shown in Table 1, the dataset is randomly divided into training, validation, and testing sets in an 8:1:1 ratio. The Chinese version includes 7452 discourses and 5742 quaternions, while the English version contains 5514 quaternions, significantly larger than existing ABSA datasets (Cai et al., 2021; Shi et al., 2022). "Dia", "Utt", and "Spk" represent dialogue, utterance, and speaker in that order. Similarly, "Tgt", "Asp", and "Opi" indicate target, aspect, and opinion terms. Each sentence averages one sentiment expression, facilitating task prediction. "Intra" and "Cross" distinguish between intra-utterance and cross-utterance quadruples. A quadruple is considered cross-utterance when any two elements (target, aspect, or opinion) are in different utterances. The dataset features about 5 speakers per conversation and includes 1275 (22.2%, Chinese) and 1227 (22.3%, English) cross-utterance level quadruples. This dataset not only provides a foundation for DiaASQ but also offers a rich resource for advancing dialogue-level sentiment analysis research.

4.2. Baselines

In this section, we review several foundational models that have significantly impacted the field of ABSA and related tasks. These models not only establish benchmarks for evaluating new approaches but also provide a framework for understanding advancements in sentiment analysis.

- **CRF-ExtractClassify** (Cai et al., 2021): This work introduces a novel task and dataset for Aspect-Category-Opinion-Sentiment (ACOS) analysis. The authors propose four baseline experiments that empirically demonstrate the efficacy of the new task in addressing implicit aspect and viewpoint challenges. The CRF-ExtractClassify model effectively integrates Conditional Random Fields (CRF) with classification techniques, thereby establishing a new standard for handling complex sentiment analysis tasks.
- **SpERT** (Markus Eberts, 2020): SpERT presents a model that utilizes attention mechanisms for the simultaneous extraction of entities and relations through spans. Notable contributions include use of lightweight reasoning on BERT embeddings, which promotes efficient recognition and filtering of entities, alongside relation classification that utilizes localized, unlabelled contextual representations. The model employs intra-sentence negative sampling within a single BERT instance, enhancing span detection and overall performance. This strategy has demonstrated high effectiveness in extracting and classifying entities and their relationships.
- **Span-ASTE** (Xu et al., 2021): Span-ASTE introduces a two-channel span pruning approach that combines guidance from both aspect term extraction (ATE) and opinion term extraction (OTE) tasks. This method enhances computational efficiency and improves the distinction between opinion and target spans. The framework demonstrates robust performance across ASTE, ATE, and OTE tasks. Additionally, its cross-level approach explicitly considers interactions between entire span components, enhancing sentiment consistency prediction through holistic semantic understanding. This model represents a significant advancement in span-based sentiment analysis.
- **paraphrase** (Zhang et al., 2021): This study introduces a novel paradigm for paraphrase modelling and presents two datasets designed to transform the Aspect Sentiment Quadruple Prediction (ASQP) task into a unified paraphrase generation process. The model effectively predicts sentiment quadruples while leveraging semantic information from natural language tags. This comprehensive generative approach minimizes the risk of error propagation commonly observed in pipeline methods and optimizes the utilization of sentiment element semantics through natural language generation. The paraphrase model provides an innovative and efficient solution for ASQP tasks.

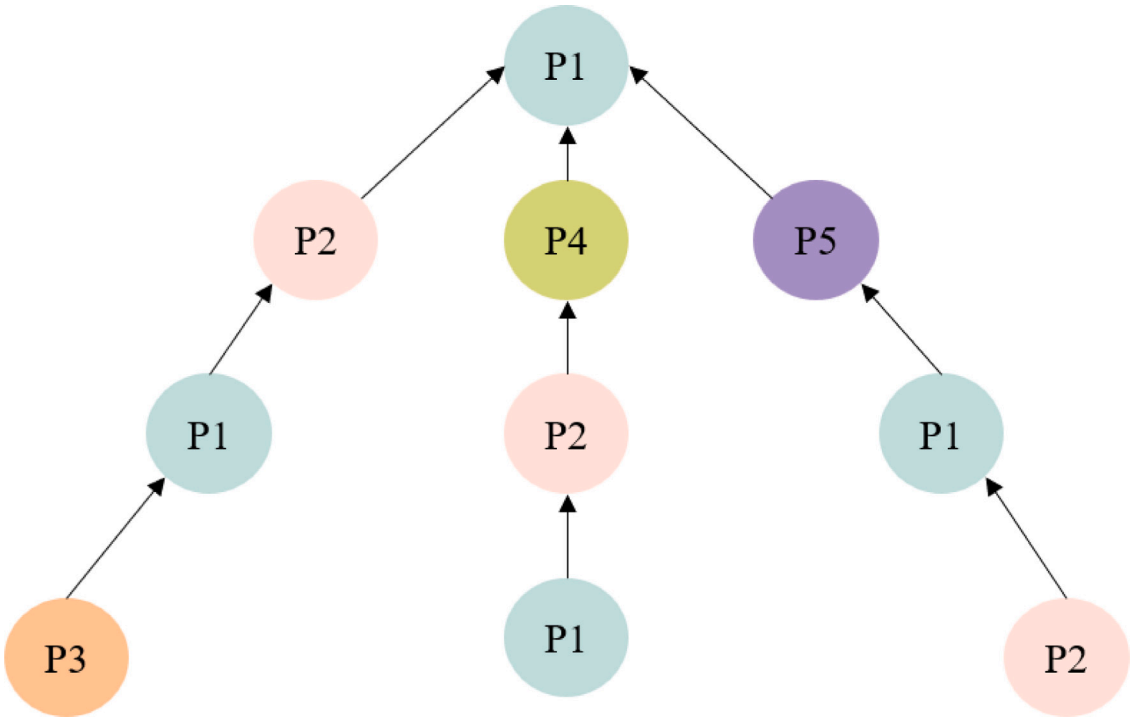


Fig. 5. The tree-like dialogue replying structure.

Table 1
Statistics for the DiaASQ dataset. Due to discrepancies in translation, the quantity of annotated items varies between the Chinese and English versions.

		Dialogue			Items			Pairs			Quadruples		
		Dia.	Utt.	Spk.	Tgt.	Asp.	Opi.	$Pair_{t-a}$	$Pair_{t-o}$	$Pair_{a-o}$	Quad.	Intra.	Cross.
ZH	Total	1000	7452	4991	8308	6572	7051	6041	7587	5358	5742	4467	1275
	Train	800	5947	3986	6652	5220	5622	4823	6062	4297	4607	3594	1013
	Valid	100	748	502	823	662	724	621	758	538	577	440	137
	Test	100	757	503	833	690	705	597	767	523	558	433	125
EN	Total	1000	7452	4991	8264	6434	6933	5894	7432	4994	5514	4287	1227
	Train	800	5947	3986	6613	5109	5523	4699	5931	3989	4414	3442	972
	Valid	100	748	502	822	644	719	603	750	509	555	423	132
	Test	100	757	503	829	681	691	592	751	496	545	422	123

- **DiaASQ** (Li et al., 2023): DiaASQ introduces the task of extracting dialogue aspect sentiment quadruple (DiaASQ) and is annotated with an extensive bilingual dataset in Chinese and English. The authors propose a baseline model that focuses on modelling relationships between word pairs, filling existing gaps in viewpoint mining and sentiment analysis within conversational contexts. This research establishes a new benchmark for sentiment analysis, expanding the potential of dialogue-based sentiment tasks.

To our knowledge, no other researchers have significantly advanced the DiaASQ task beyond the contributions of Li et al. (2023). We continue to utilize the benchmark models from this work, which are crucial for ABSA research. These benchmarks offer unique insights and methods, representing the cutting edge of the field. Comparing our model, IFusionQuad, with these benchmarks validates its effectiveness and situates our work within the broader context of ABSA's evolution.

4.3. Experiment setting details

This section outlines the technical specifications and configurations utilized throughout the experimental phase of our research.

- **Encoder and Optimizer Configuration**: As the foundation of our model, we choose Chinese-Roberta-wwm-base (Cui et al., 2021) as the base encoder for the Chinese dataset and Roberta-Large (Liu et al., 2019) for the English dataset. These models are chosen due to their proven effectiveness in capturing nuanced language representations and their state-of-the-art performance on various NLP tasks. The learning rate for BERT fine-tuning is meticulously set to 1×10^{-5} to ensure stable and gradual convergence, preventing potential overfitting and ensuring fine-grained adjustments to the pre-trained weights.
- **Learning Rate and Dropout Configuration**: We implement a dropout rate of 0.2 on the output representations of BERT to mitigate overfitting by randomly omitting units during training, which helps improve the model's generalization capability. For other trainable parameters within our model, we utilize a learning rate of 1×10^{-3}. This higher learning rate facilitates faster convergence for these parameters, which are less sensitive to fine-tuning compared to the pre-trained BERT layers.
- **Model Architecture Specifications**: The testing outcomes are derived from models optimized on the development set. To ensure the robustness and reliability of the results, all experiments were conducted using two distinct random seeds, with the final scores calculated as the average across the two runs. This practice addresses variability due to random initialization and training stochasticity.
- **Training Configuration**: The model is trained for 10 epochs, a duration selected to strike a balance between adequate training

Table 2
Results of the DiaASQ task for both the Chinese and English datasets. 'T, A, O' stands for Target, Aspect, and Opinion, respectively.

		Span Match (F1)			Pair Extraction (F1)			Quadruple (F1)	
		T	A	O	T-A	T-O	A-O	Micro	Iden.
ZH	CRF-Extract-Classify	91.11	75.24	50.06	32.47	26.78	18.90	8.81	9.25
	SpERT	90.69	76.81	54.06	38.05	31.28	21.89	13.00	14.19
	ParaPhrase	/	/	/	37.81	34.32	27.76	23.27	27.98
	Span-ASTE	/	/	/	44.13	34.46	32.21	27.42	30.85
	DiaASQ	90.23	**76.94**	59.35	48.61	43.31	45.44	34.94	37.51
	Our IFusionQuad	**91.69**	75.9	**60.96**	**54.68**	**51.81**	**50.04**	**41.53**	**44.56**
EN	CRF-Extract-Classify	88.31	71.71	47.9	34.31	20.94	19.21	11.59	12.80
	SpERT	87.82	74.65	54.17	28.33	21.39	23.64	13.07	13.38
	ParaPhrase	/	/	/	37.22	32.19	30.78	24.54	26.76
	Span-ASTE	/	/	/	42.19	30.44	45.90	26.99	28.34
	DiaASQ	**88.62**	**74.71**	60.22	47.91	45.58	44.27	33.31	36.80
	Our IFusionQuad	88.31	74.23	**63.48**	**52.65**	**51.82**	**51.94**	**35.96**	**41.49**

and avoiding overfitting. The training process is performed using an RTX3070 graphics card, with the batch size set to 1 due to resource limitations. This configuration ensures that the model can be trained effectively within the computational constraints, while still providing sufficient exposure to the training data.

- **Evaluation Metrics**: We adopt the F1 metric for evaluation. For span prediction, correct matches must align with both left and right boundaries; for pair prediction, matches should align with both spans and the relation; for quad prediction, matches should align with all four elements precisely. Our primary focus lies on quadruple extraction performance. Accordingly, we employ micro F1 and identification F1 metrics, with micro F1 evaluating the entire quad, while identification F1 (Barnes et al., 2021) does not differentiate polarity.

The implementation details of the IFusionQuad model are carefully crafted to ensure that the experimental results are both reliable and truly reflective of the model's capabilities in addressing the DiaASQ task. This meticulous attention to detail ensures robustness and accuracy in the evaluation, showcasing the model's effectiveness and potential in solving the target problem.

4.4. Main results

This subsection presents the primary outcomes derived from our experiments on the DiaASQ tasks, as summarized in Table 2. The analysis of the results indicates that our proposed approach consistently achieves the highest performance across nearly all metrics. A notable highlight is the outstanding performance of the IFusionQuad model.

First, the performance difference among models in span detection is minimal, with all methods performing well in this subtask. This may be attributed to the relative simplicity of recognizing mentions without considering the interrelationships between term types (T/A/O). Second, our model significantly surpasses the baseline in pairwise detection, excelling in the T-A, T-O, and A-O tasks. This indicates our model's enhanced capability to extract emotional information within conversational contexts. Lastly, substantial improvements are observed in quaternion extraction. For Chinese extraction, our model shows a 6.59% increase in micro F1 (41.53 vs. 34.94) and a 7.05% increase in identification F1 (44.56 vs. 37.51). For the English dataset, micro F1 improved by 2.65% (35.96 vs. 33.31), and identification F1 increased by 4.69% (41.49 vs. 36.80).

Overall, IFusionQuad consistently outperformed the baseline models across all evaluation datasets, these results clearly demonstrate the effectiveness of our model in the DiaASQ task, validating its SOTA status.

5. Model analysis

This section offers a comprehensive analysis of the IFusionQuad model, focusing on its innovative components and their contributions to enhancing performance in the DiaASQ task. We start with an ablation study that evaluates the individual impact of each module on the model's overall effectiveness. Next, we delve into the specifics of the CloBlock module, examining its role in refining the model's ability to discern local and global features within dialogues. We also investigate the performance of cross-utterance quadruple extraction, showcasing the model's robustness in handling complex interactions. Furthermore, we present two detailed case studies that illustrate the model's capabilities and identify areas for further improvement. This thorough analysis underscores the model's strengths and highlights its potential to advance dialogue-based sentiment analysis. Finally, we conduct two additional analyses: one comparing our model's performance with large language models, and another evaluating its robustness and scalability.

5.1. Ablation study

The IFusionQuad model integrates multiple innovative cutting-edge modules aimed at boosting its effectiveness in the DiaASQ task. Table 3 provides a comprehensive analysis of various ablation studies and assesses the impact of individual components on the model's overall performance. The table is divided into two main sections corresponding to the languages ZH (Chinese) and EN (English), with performance metrics evaluated using Micro F1, Intra-Utt., and Inter-Utt. scores. To evaluate the contributions of various components, we performed a series of ablation experiments. By systematically removing key elements of the model and monitoring the subsequent decline in performance, we were able to assess their individual impact.

- **Without All-InteractiveNet Fusion**: Removing this component causes a significant decrease in performance, with the Micro F1 score dropping by 6.88 points for ZH and 3.65 points for EN. The reductions in Intra-Utt. and Inter-Utt. scores are also substantial, highlighting the critical role of interactive fusion in our model.
- **Without CloBlock**: The absence of CloBlock leads to a Micro F1 score decrease of 6.20 for ZH and 2.41 for EN. This indicates that CloBlock plays an essential role in maintaining the model's contextual understanding, particularly in the ZH dataset where the drop in Intra-Utt. and Inter-Utt. scores are more pronounced.
- **Without Gating Mechanism**: The gating mechanism is also pivotal, as its absence leads to a minor reduction in performance metrics. The Micro F1 score decreases by 0.66 for ZH and 1.11 for EN, showing that while the impact is less severe compared to other components, the gating mechanism still contributes to the model's overall efficacy.

Table 3
Ablation results (Micro F1). 'w/o All-InteractiveNet Fusion': removing all InteractiveNet Fusion items.

	ZH			EN		
	Micro F1	Intra-Utt.	Inter-Utt.	Micro F1	Intra-Utt.	Inter-Utt.
Our IFusionQuad	41.53	44.25	30.21	35.96	39.84	19.78
w/o All-InteractiveNet Fusion	34.65 (↓6.88)	37.10 (↓7.15)	25.30 (↓4.91)	32.31 (↓3.65)	34.40 (↓5.44)	16.79 (↓2.99)
w/o CloBlock	35.33 (↓6.20)	38.50 (↓5.75)	23.79 (↓6.42)	33.55 (↓2.41)	37.72 (↓2.12)	17.66 (↓2.12)
w/o Gating Mechanism	40.87 (↓0.66)	43.69 (↓0.56)	28.86 (↓1.35)	34.85 (↓1.11)	38.35 (↓1.49)	18.75 (↓1.03)
w/o Biaffine Attention	37.17 (↓4.36)	40.56 (↓3.69)	26.24 (↓3.97)	33.84 (↓2.12)	36.06 (↓3.78)	18.45 (↓1.33)

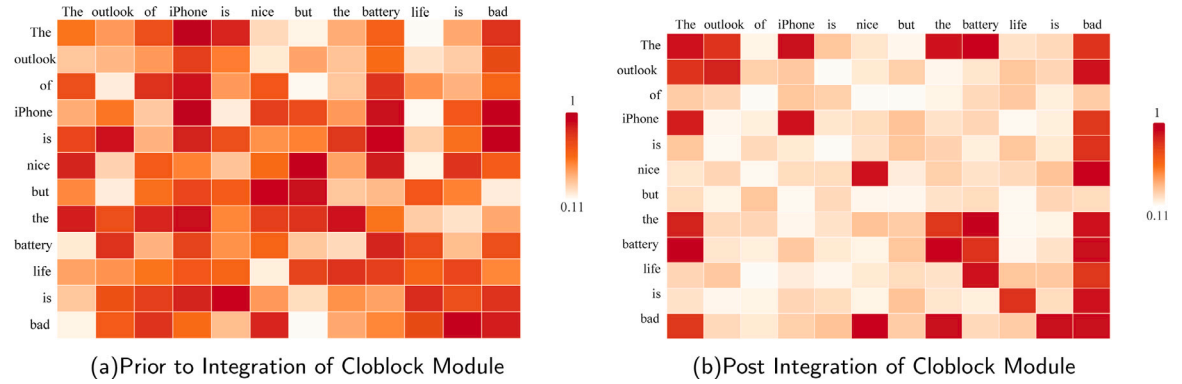


Fig. 6. Visualizing the Impact of the Cloblock Module on an Example using Heatmaps.

- **Without Biaffine attention**: Eliminating the Biaffine attention mechanism leads to a reduction in the Micro F1 score by 4.36 for ZH and 2.12 for EN. The decrease in performance metrics suggests that Biaffine attention is crucial for capturing fine-grained interactions within the data.

The analysis presented in Table 3 underscores the critical role of each component in the IFusionQuad model. Notably, CloBlock markedly improves the model's semantic understanding and prediction. The performance decline with the removal of individual components highlights the synergistic benefits of the integrated architecture, validating the holistic design of our approach.

5.2. CloBlock module

The focused approach facilitated by the Cloblock module improves the capacity of the model to identify both local and global features. To assess its efficacy, we utilized a pertinent example: "The outlook of the iPhone is nice but the battery life is bad", which encapsulates two distinct yet interconnected aspects: appearance and battery life. Our experimental methodology was devised to scrutinize the impact of the Cloblock module on the model's performance. Employing heatmaps as Fig. 6, we visualized the model's attention distribution during sentence processing. Before utilizing the Cloblock module, attention was uniformly distributed across the entire sentence, resulting in a lack of discernible focal points. Conversely, the integration of the Cloblock module led to a heightened focus on specific targets such as "iPhone", along with particular aspects like "outlook" and "battery life", and sentiment words like "nice" and "bad" within the sentence. This observation underscores the Cloblock module's effectiveness in enhancing the model's perceptual acuity, facilitating a more nuanced understanding of the relative importance of different sentence components.

5.3. Cross-utterance quadruple extraction

This subsection evaluates the performance of cross-utterance quadruple extraction to directly study the efficacy of our proposed model. We conducted experiments at various cross-utterance levels to evaluate the effect of our innovative components on the model's performance. As depicted in Fig. 7, the red line represents our IFusionQuad model, while the blue and pink lines indicate two baseline models. The orange line illustrates the results obtained after removing the InteractiveNet Fusion component from our model. Notably, when the cross-utterance level reaches three or more (⩾3), the baseline systems face significant challenges, frequently failing to recognize any quadruples. This failure highlights a substantial limitation in the baseline models' ability to manage complex, multi-utterance dependencies. In contrast, our IFusionQuad model effectively tackles this challenge, demonstrating robust performance even at higher cross-utterance levels (⩾3). The data in Fig. 7 clearly demonstrates the superiority of the IFusionQuad model in extracting cross-utterance quadruples.

5.4. Case study

5.4.1. Case 1

This section provides an in-depth case study to illustrate the capabilities of our model in Quadruple extraction for the DiaASQ task. The case study is visualized through word clouds and network graphs, as shown in Figs. 8(a) and 8(b), respectively.

Fig. 8(a) presents a word cloud visualization of the dialogue data which is sourced from the test set of the English version of the datasets mentioned in Section 4.1. In this visualization, words with higher frequency are represented in larger fonts. Upon closer examination, it becomes apparent that words associated with targets, aspects, and opinions appear infrequently within the dialogue data. This observation reflects real-life scenarios where such terms are not always prominently featured in conversations, underscoring the challenge of extracting meaningful information from sparse data. Fig. 8(b) displays a network graph that visualizes the extracted quadruples. The network graph data is derived from a carefully selected sample within the same test set to facilitate clear visualization. This sample was chosen to represent the various aspects of the dialogue data and to showcase the efficacy of our model in a comprehensible manner. In this graph, grey nodes represent the two primary target entities, 'Samsung' and 'iPhone', which are positioned centrally, indicating their significance in the analysis. Connected to these targets are various aspects associated with each product, depicted as green nodes. For instance, Samsung's aspects

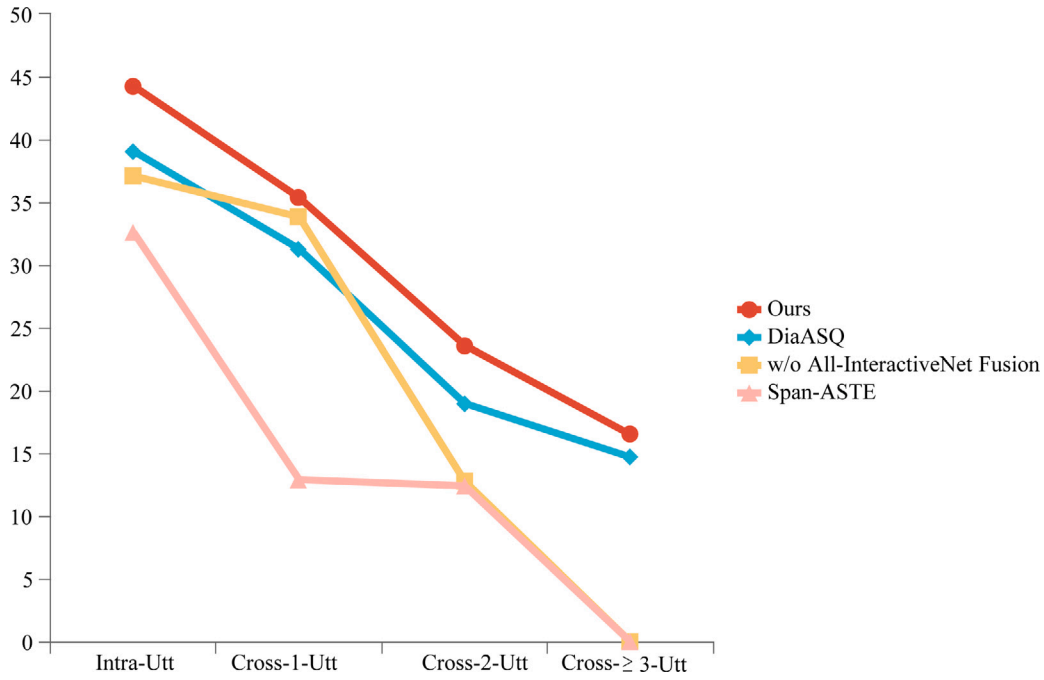


Fig. 7. Results Across Various Cross-Utterance Levels.

include 'experience', 'notification', 'one-UI function', and 'humanization', while iPhone's aspects encompass 'appearance', 'photography', and 'charging'. The sentiment words associated with specific aspects are represented by red and blue nodes, indicating the nature of the feedback. Red nodes signify positive sentiments, while blue nodes signify negative sentiments. This colour-coding provides a clear visual distinction between different types of feedback. The network graph reveals that Samsung is frequently associated with aspects such as 'notification information' and 'humanization'. The sentiments linked to these aspects range from positive (e.g., 'enjoyment') to negative (e.g., 'could not be folded'). Conversely, feedback for the iPhone is more concentrated on 'system experience' and 'appearance', with notable negative sentiments such as 'worst' and 'slow'.

This case study highlights our model's ability to accurately analyse real-life sentences and extract meaningful quadruples. The visualizations provide valuable insights into the advantages and limitations of various products, which can be leveraged for product improvement and marketing strategies. The clear distinction between targets, aspects, and sentiments in the network graph emphasizes the effectiveness of our model in handling complex dialogue data, thus providing a comprehensive understanding of user feedback.

5.4.2. Case 2

To better understand the model's strengths and weaknesses, we analyse its errors and identify areas for improvement. This case study utilizes data from the English version of the DiaASQ test set. We tested the entire English test set using our model and randomly selected a misclassified example for detailed analysis. Fig. 9(a) shows the dialogue data "This phone is not... who uses it, who knows".; Fig. 9(b) displays the corresponding gold label. Fig. 10 illustrates the model's extracted output. To pinpoint where errors occur, we compare the Target, Aspect, Opinion, and Quadruple elements. The model correctly identifies Target and Aspect but makes errors in Opinion extraction, leading to incorrect Quadruple formation. The following examples demonstrate sentiment polarity misjudgments:

- "iPhone, experience, better, pos": This should be negative (neg) but was misjudged as positive (pos).
- "Mi 11, experience, speechless, pos": This should be negative (neg) but was misjudged as positive (pos).

The sentiment polarity errors in quadruple generation can be traced back to inaccuracies in Opinion extraction, which highlight difficulties in identifying emotion words associated with the Target and Aspect, along with their corresponding polarity. Potential causes for these Opinion extraction errors include:

1. **Misinterpretation of polysemous words:**
   Words like "better" typically convey positive sentiment but can imply negativity in contrastive contexts, such as in ironic statements. Misunderstanding such nuances can lead to misclassification.
2. **Contextual misclassification:**
   Words like "speechless" can express both positive and negative emotions depending on the context. Failure to capture the context accurately may result in incorrect sentiment classification.
3. **Attention mechanism bias:**
   During the InteractiveNetFusion phase, incorrect attention focus on seemingly positive words (e.g., "better") can cause the model to overlook sarcasm or contrasting sentiments, leading to sentiment polarity errors.

This case study provides insights into the specific steps where errors occur, demonstrating the model's limitations. Visualizations have highlighted the root causes of these errors, offering valuable guidance for future improvements.

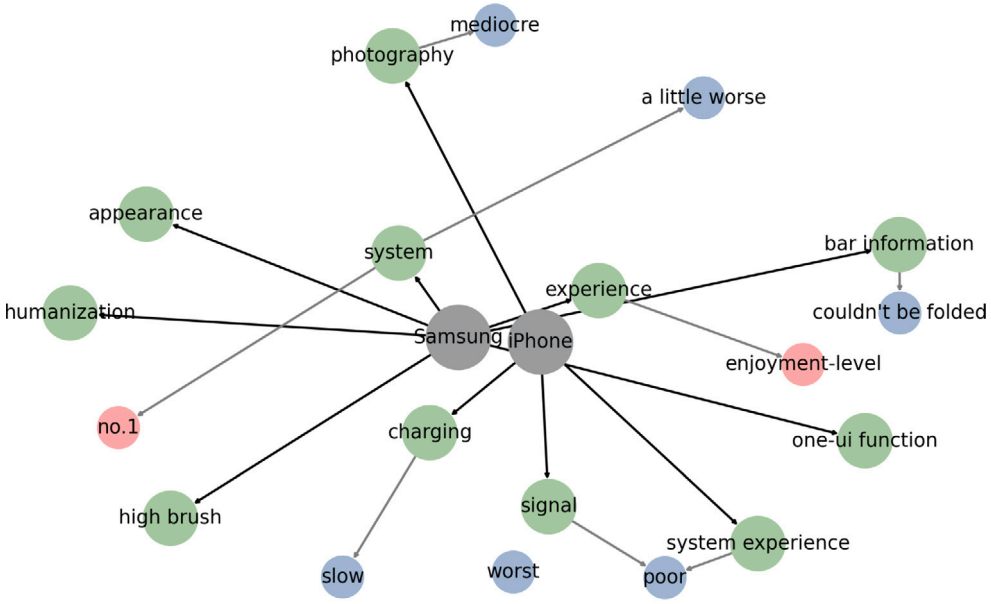
5.5. Further analysis

In this section, we consider diving into the model performances and carry on in-depth analyses to better understand the strengths of our method.

5.5.1. Comparing with large language models

Considering the popularity of large language models (LLMs), as illustrated in Table 4 we conducted experiments using several open-source large language models, including Qwen1.5-1.8B-Chat, ChatGLM3-6B-Base, and Qwen-14B-Chat-Int4. We used the Instruction

(a) Visual Word Cloud for a Dialogue



(b) Network Relationship Diagram of Extracted Quadruples

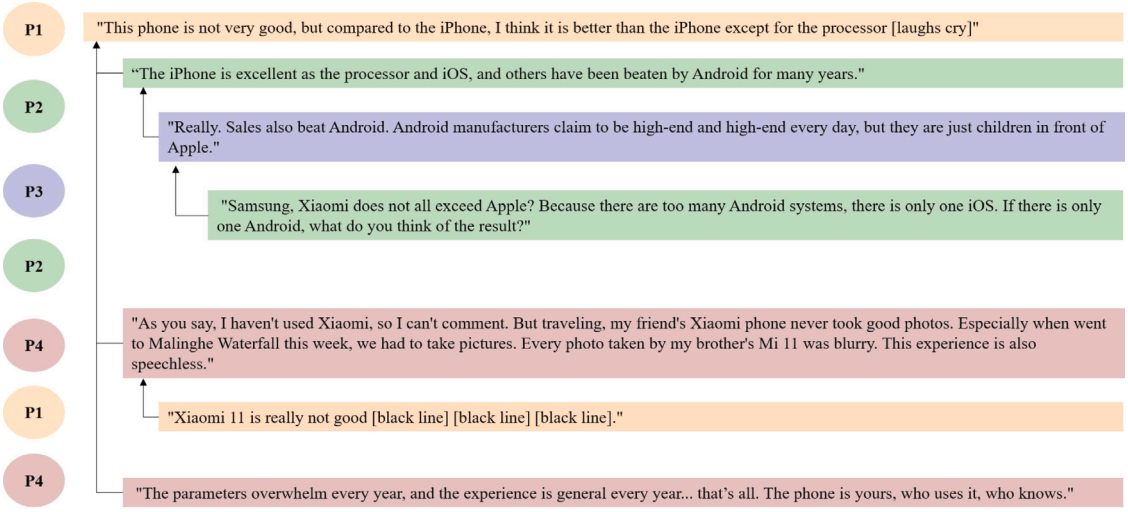
Fig. 8. Visualization of a Dialogue Case.

Tuning approach (Varia et al., 2023) for these open-source LLMs in the DiaASQ task and tested these models on both Chinese (ZH) and English (EN) datasets to provide a comprehensive comparison.

The results indicate a general trend of performance improvement in LLMs as the parameter size increases in the DiaASQ task. For instance, in the ZH dataset, the Qwen-14B-Chat-Int4 model, which has the largest parameter size among the compared models, achieved the highest performance, with a quadruple F1 score of 42.27% in the Micro metric and 42.45% in the Identification (Iden.) metric. However, our proposed IFusionQuad model outperformed Qwen-14B-Chat-Int4 in the Iden. metric, achieving a quadruple F1 score of 44.56%, which is 2.11% higher. In the EN dataset, Qwen-14B-Chat-Int4 attained a quadruple F1 score of 35.45% in the Micro metric and 40.12% in the Macro metric. Nevertheless, our IFusionQuad model exceeded these scores, obtaining 35.96% in the Micro metric and 41.49% in the Iden. metric, surpassing Qwen-14B-Chat-Int4 by 0.51% and 1.37%, respectively. Moreover, our model consistently outperformed three other smaller LLMs in the DiaASQ task. It is evident that while LLMs are capable of executing the DiaASQ task, their effectiveness remains limited. Even LLMs with over 10B parameters, such as Qwen-14B-Chat-Int4, demonstrated strong performance on both the ZH and EN datasets. However, our IFusionQuad model not only matches these results but also surpasses them in certain instances. This highlights the effectiveness of our approach, indicating that IFusionQuad can achieve competitive, and at times superior, results even in the presence of very powerful LLMs.

5.5.2. *Evaluating model robustness and scalability*

Considering that there is only one publicly available dataset for the DiaASQ task and almost no other relevant research on the task, to ensure that our model has not only excellent performance in the DiaASQ task but also has good task scalability and robustness. We tested it on the Aspect-Category-Opinion-Sentiment Quadruple Extraction (ASQE)



(a) A Dialogue Instance from DiaASQ Dataset's English Testing Set

Gold Label

Targets: "iPhone", "iPhone", "Samsung", "Xiaomi", "Mi 11", "Xiaomi 11", "iPhone", "iPhone"

Aspects: "Sales", "experience", "experience", "The parameters", "processor", "processor", "iOS", "experience", "image system", "photos"

Opinions: "better, pos", "overwhelm, pos", "beat, pos", "not good, neg", "blurry, neg", "excellent, pos", "speechless, neg", "exceed, pos", "general, neg", "backward, neg", "better, neg"

Quadruples:

"iPhone, processor, better, pos", "iPhone, processor, excellent, pos",
"iPhone, iOS, excellent, pos ", "iPhone, Sales, beat, pos",
"Samsung, Sales, exceed, pos ", "Xiaomi, Sales, exceed, pos",
"Xiaomi 11, photos, not good, neg " "iPhone, image system, backward, neg"
"iPhone, experience, better, neg ", "Mi 11, photos, blurry, neg",
" Mi 11, experience, speechless, neg "

(b) The Gold Label For The Dialogue Instance

Fig. 9. Instance from DiaASQ Dataset's English Testing Set.

Table 4
Results of the DiaASQ task for large language models.

Dataset	Model	Quadruple F1	
		Micro	Iden.
ZH	Qwen1.5-1.8B-Chat	27.89	30.01
	ChatGLM3-6B-Base	40.21	41.12
	Qwen-14B-Chat-Int4	**42.27**	42.45
	Our IFusionQuad	41.53	**44.56**
EN	Qwen1.5-1.8B-Chat	24.58	26.23
	ChatGLM3-6B-Base	33.29	38.67
	Qwen-14B-Chat-Int4	35.45	40.12
	Our IFusionQuad	**35.96**	**41.49**

task and compared it with previous state-of-the-art models. There are well-established studies on this task and two classical public datasets, Restaurant-ACOS and Laptop-ACOS, which were meticulously curated by Cai et al. (2021) as shown in Table 5. This table presents a summary of the statistical characteristics of the Restaurant-ACOS and Laptop-ACOS datasets. The table is structured to provide insights into the composition of the datasets, including the total number of sentences, and the distribution of quadruples across various categories.

As shown in Table 6, we compare our model's performance on the ASQE task with previous state-of-the-art models using the Restaurant-ACOS and Laptop-ACOS datasets. textbfBARTABSA (Yan et al., 2021) ingeniously reformulates all aspect-based sentiment analysis sub-tasks into a unified generation task, with a focus on generating category indices. This model has been pivotal in showcasing the versatility and potential of generative models in sentiment analysis. GAS (Zhang et al., 2021c) proposes a unified generation framework for ABSA tasks, conceptualizing them as challenges in sentiment element sequence generation. This innovative approach has broadened the applicability of generation models in sentiment analysis. **Paraphrase** (Zhang et al., 2021) presents a novel paradigm for paraphrase modelling and introduces two datasets aimed at converting the Aspect Sentiment Quadruple Prediction (ASQP) task into a unified paraphrase generation process. **Opinion Tree Generation** (Bao et al., 2022) envisions a tree-structured semantic representation for the joint detection of sentiment elements, offering a nuanced representation of sentiment structure. **EMRC** (Ye et al., 2023) adopts a multi-turn machine reading comprehension (MRC) approach to the ASQE task, showcasing the

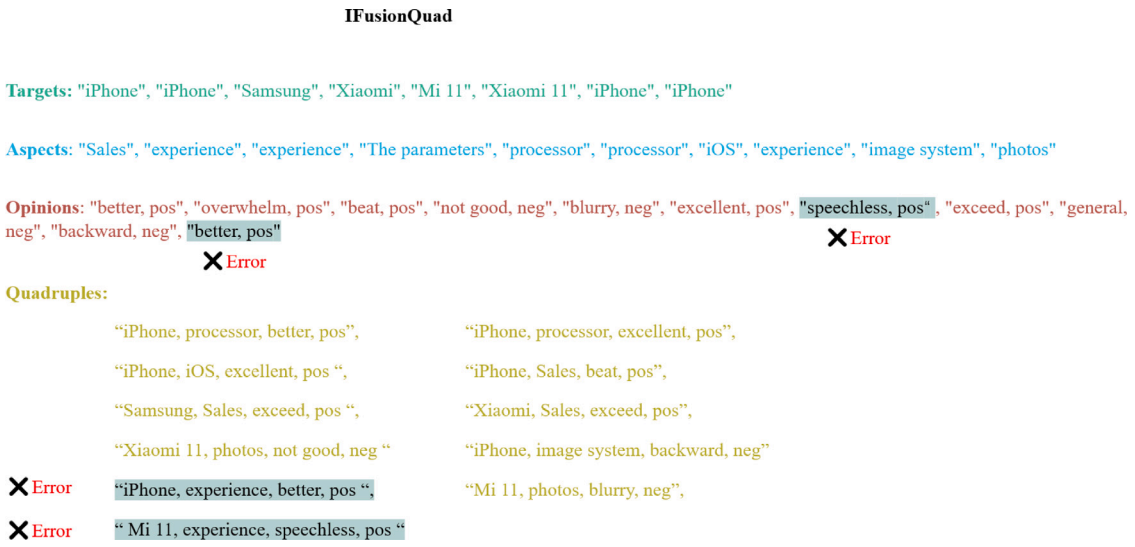


Fig. 10. Information Taken By Our IFusionQuad. Errors In Comparison With Gold Labels Are Highlighted And Annotated.

Table 5
Summary of Experimental Dataset Statistics: The dataset is categorized into four types of terms: 'EA' for explicit aspect terms, 'EO' for explicit opinion terms, 'IA' for implicit aspect terms, and 'IO' for implicit opinion terms. Each category is quantitatively represented to facilitate comprehensive analysis.

	Quadruplets						$\frac{Quadruplets}{Sentences}$	Categories
	Sentences	EA&EO	EA&IO	IA&EO	IA&IO	All		
Restaurant-ACOS	2286	2429 (66.40%)	350 (9.57%)	530 (14.49%)	349 (9.54%)	3658	1.60	13
Laptop-ACOS	4076	3269 (56.77%)	1237 (21.48%)	910 (15.80%)	342 (5.94%)	5758	1.42	121

Table 6
Performance comparison of our model and State-of-the-Art models on Restaurant-ACOS and Laptop-ACOS datasets.

METHOD	**Restaurant-ACOS**			**Laptop-ACOS**		
	P	R	F1	P	R	F1
BARTABSA	56.62	55.35	55.98	41.65	40.46	41.05
GAS	60.69	58.52	59.59	41.6	42.75	42.17
Paraphrase	58.98	59.11	59.04	41.77	**45.04**	43.34
Opinion tree generation	63.96	61.74	62.83	46.11	44.79	45.44
EMRC	**64.97**	61.18	63.02	47.27	44.66	**45.92**
Our Model	64.12	**62.33**	**63.81**	**47.78**	44.77	45.89

versatility of MRC methodologies in tackling complex extraction tasks. This method has underscored the potential of adapting MRC techniques for sentiment analysis.

As seen from Table 6, our model shows strong performance across both datasets. Specifically, on the Restaurant-ACOS dataset, our model achieves a Precision (P) of 64.12, Recall (R) of 62.33, and an F1 score of 63.81. These results are competitive with the current best-performing model (EMRC), even surpassing it in terms of Recall and F1. On the Laptop-ACOS dataset, our model achieves the highest Precision at 47.78, with an F1 score of 45.89, which is very close to the top F1 score of 45.92 achieved by the EMRC model.

Overall, these results demonstrate that our model performs comparably to other state-of-the-art models and, in some cases, outperforms them in key metrics. This highlights the robustness and scalability of our model across different tasks.

6. Conclusions

This study presents the IFusionQuad model, a pioneering end-to-end framework for dialogue-based aspect sentiment quadruple extraction (DiaASQ). IFusionQuad effectively enhances local feature perception, enabling the model to capture both high and low-frequency information. The framework incorporates a range of innovative components that synergistically improve its ability to detect and interpret nuanced sentiment expressions in conversational contexts.

Specifically, IFusionQuad addresses critical limitations of existing state-of-the-art methods in DiaASQ tasks by providing solutions to key challenges:

1. Improved Long-Distance Dependency Handling: Traditional methods often struggle to capture long-distance dependencies and conversational context across multiple dialogue turns. IFusionQuad leverages RoPE to reduce computational complexity while enhancing the model's capacity to manage these dependencies, surpassing the limitations of both graph-based and sequence-to-sequence methods.
2. Efficient Local and Global Feature Extraction: Graph-based methods typically encounter computational inefficiencies and fail to simultaneously capture local and global features. Sequence-to-sequence methods also face challenges in managing contextual nuances. IFusionQuad integrates the CloBlock module, which employs convolution, pooling, and gating mechanisms to enhance the extraction of both high and low-frequency features, thereby improving the model's capacity to understand complex dialogue contexts.

3. Noise Reduction and Exposure Bias Mitigation: Existing approaches can be affected by noise from informal language and incomplete sentences, while sequence-to-sequence methods often suffer from exposure bias. IFusionQuad effectively mitigates these issues through the CloBlock module and Biaffine attention, filtering out noise and focusing on critical conversational features. This leads to more accurate quadruple extraction, addressing the noise and bias inherent in previous methods.
4. Effective Fusion of Dialogue-Specific Features: Many current methods underutilize dialogue-specific structures such as speaker identity and dialogue threads. IFusionQuad's design integrates these features effectively, enabling a better understanding and extraction of nuanced dialogue elements.

Furthermore, by addressing specific limitations of existing methods and integrating advanced features, IFusionQuad offers a more robust, scalable, and accurate solution for dialogue-based sentiment analysis.

Our comprehensive testing on benchmark datasets provides empirical validation demonstrating the superior performance of the IFusionQuad model compared to baseline models. Extensive comparative analyses indicate that IFusionQuad achieves state-of-the-art results in dialogue understanding and sentiment analysis. Specifically, the model achieved a 6.59% increase in micro F1 and a 7.05% increase in identification F1 for Chinese extraction, while the English dataset witnessed a 2.65% improvement in micro F1 and a 4.69% enhancement in identification F1.

The IFusionQuad model stands as a testament to the power of innovative framework design and methodological integration in advancing the field of dialogue understanding and sentiment analysis. Our findings contribute to both the theoretical understanding of these domains and practical implications for real-world applications requiring a nuanced grasp of dialogue and sentiment dynamics.

CRediT authorship contribution statement

Haoyu Jiang: Conceptualization, Data curation, Formal analysis, Visualization, Writing – original draft. **Xiaoliang Chen:** Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **Duoqian Miao:** Funding acquisition. **Hongyun Zhang:** Validation. **Xiaolin Qin:** Funding acquisition, Resources. **Xu Gu:** Software. **Peng Lu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by the Science and Technology Program of Sichuan Province (Grant no. 2023YFS0424), the National Key R&D Plan "Key Special Project of Cyberspace Security Governance" (No. 2022YFB3104700), the National Natural Science Foundation (Grant nos. 61976158, 62376198), the Science and Technology Service Network Initiative (No. KFJ-STS-QYZD-2021-21-001), and the Talents by Sichuan provincial Party Committee Organization Department, and Chengdu - Chinese Academy of Sciences Science and Technology Cooperation Fund Project (Major Scientific and Technological Innovation Projects)

Data availability

Data and code are available for download at the following web links. https://github.com/Joeisjoejoe/IFusionQuad.

References

Bao, X., Zhongqing, W., Jiang, X., Xiao, R., & Li, S. (2022). Aspect-based sentiment analysis with opinion tree generation. In *Proceedings of the thirty-first international joint conference on artificial intelligence, IJCAI-22* (pp. 4044–4050). International Joint Conferences on Artificial Intelligence Organization, http://dx.doi.org/10.24963/ijcai.2022/561.

Barnes, J., Kurtz, R., Oepen, S., Øvrelid, L., & Velldal, E. (2021). Structured sentiment analysis as dependency graph parsing. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing* (pp. 3387–3402). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.263.

Cai, H., Xia, R., & Yu, J. (2021). Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.29.

Chen, H., Zhai, Z., Feng, F., Li, R., & Wang, X. (2022). Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th annual meeting of the association for computational linguistics*. http://dx.doi.org/10.18653/v1/2022.acl-long.212.

Consoli, S., Barbaglia, L., & Manzan, S. (2022). Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. *Knowledge-Based Systems*, *247*, Article 108781. http://dx.doi.org/10.1016/j.knosys.2022.108781.

Cui, Y., Che, W., Liu, T., Qin, B., & Yang, Z. (2021). Pre-training with whole word masking for Chinese BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 3504–3514. http://dx.doi.org/10.1109/TASLP.2021.3124365.

Dai, J., Yan, H., Sun, T., Liu, P., & Qiu, X. (2021). Does syntax matter? A strong baseline for aspect-based sentiment analysis with RoBERTa. In *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: human language technologies*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.naacl-main.146.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *North American chapter of the association for computational linguistics*. https://api.semanticscholar.org/CorpusID:52967399.

Dozat, T., & Manning, C. D. (2016). Deep biaffine attention for neural dependency parsing. http://dx.doi.org/10.48550/arXiv.1611.01734, arXiv preprint arXiv:1611.01734.

Fan, Q., Huang, H., Guan, J., & He, R. (2023). Rethinking local perception in lightweight vision transformer. arXiv:2303.17803, https://api.semanticscholar.org/CorpusID:257901238.

Fu, Y., Yuan, S., Zhang, C., & Cao, J. (2023). Emotion recognition in conversations: A survey focusing on context, speaker dependencies, and fusion methods. *Electronics*, *12*(22), http://dx.doi.org/10.3390/electronics12224714.

Ghosal, D., Majumder, N., Poria, S., Chhaya, N., & Gelbukh, A. (2019). Dialoguegcn: A graph convolutional neural network for emotion recognition in conversation. In *Conference on empirical methods in natural language processing*. https://api.semanticscholar.org/CorpusID:201698197.

Gui, T., Zhang, Q., Huang, H., Peng, M., & Huang, X. (2017). Part-of-speech tagging for Twitter with adversarial neural networks. In *Conference on empirical methods in natural language processing*. http://dx.doi.org/10.18653/v1/D17-1256.

He, K., Mao, R., Gong, T., Li, C., & Cambria, E. (2023). Meta-based self-training and re-weighting for aspect-based sentiment analysis. *IEEE Transactions on Affective Computing*, *14*(3), 1731–1742. http://dx.doi.org/10.1109/TAFFC.2022.3202831.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. http://dx.doi.org/10.1162/neco.1997.9.8.1735.

Jbene, M., raif, M., Tigani, S., Chehri, A., & Saadane, R. (2022). User sentiment analysis in conversational systems based on augmentation and attention-based bilstm. *Procedia Computer Science*, *207*, 4106–4112. http://dx.doi.org/10.1016/j.procs.2022.09.473.

Kiritchenko, S., Zhu, X., Cherry, C., & Mohammad, S. (2014). NRC-Canada-2014: Detecting aspects and sentiment in customer reviews. In *Proceedings of the 8th international workshop on semantic evaluation (semEval 2014)*. http://dx.doi.org/10.3115/v1/S14-2076.

Li, X., Bing, L., Li, P., & Lam, W. (2019). A unified model for opinion target extraction and target sentiment prediction. AAAI Press, http://dx.doi.org/10.1609/aaai.v33i01.33016714.

Li, X., Bing, L., Zhang, W., & Lam, W. (2019). Exploiting BERT for end-to-end aspect-based sentiment analysis. In *Proceedings of the 5th workshop on noisy user-generated text (W-NUT 2019)* (pp. 34–41). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/D19-5505.

Li, B., Fei, H., Li, F., Wu, Y., Zhang, J., Wu, S., Li, J., Liu, Y., Liao, L., Chua, T.-S., & Ji, D. (2023). Diaasq : A benchmark of conversational aspect-based sentiment quadruple analysis. http://dx.doi.org/10.48550/arXiv.2211.05705.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. arXiv:1907.11692, https://api.semanticscholar.org/CorpusID:198953378.

Liu, Z., Xia, R., & Yu, J. (2021). Comparative opinion quintuple extraction from product reviews. In *Proceedings of the 2021 conference on empirical methods in natural language processing*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.322.

Liu, S., Zhou, J., Zhu, Q., Chen, Q., Bai, Q., Xiao, J., & He, L. (2024). Let's rectify step by step: Improving aspect-based sentiment analysis with diffusion models. In *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (LREC-COLING 2024)* (pp. 10324–10335). Torino, Italia: ELRA and ICCL, https://aclanthology.org/2024.lrec-main.902.

Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., & Cambria, E. (2018). Dialoguernn: An attentive RNN for emotion detection in conversations. In *AAAI conference on artificial intelligence*. https://api.semanticscholar.org/CorpusID:53172956.

Mao, R., Liu, Q., He, K., Li, W., & Cambria, E. (2023). The biases of pre-trained language models: An empirical study on prompt-based sentiment analysis and emotion detection. *IEEE Transactions on Affective Computing*, *14*(3), 1743–1753. http://dx.doi.org/10.1109/TAFFC.2022.3204972.

Mao, Y., Shen, Y., Yu, C., & Cai, L. (2021). A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*. http://dx.doi.org/10.1609/aaai.v35i15.17597.

Markus Eberts, A. U. (2020). Span-based joint entity and relation extraction with transformer pre-training. *ECAI*, *325*, http://dx.doi.org/10.48550/arXiv.1909.07755.

Mukherjee, R., Nayak, T., Butala, Y., Bhattacharya, S., & Goyal, P. (2021). PASTE: A tagging-free decoding framework using pointer networks for aspect sentiment triplet extraction. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 9279–9291). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.731.

Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., & Si, L. (2020). Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence*. http://dx.doi.org/10.1609/aaai.v34i05.6383.

Phan, M. H., & Ogunbona, P. O. (2020). Modelling context and syntactical features for aspect-based sentiment analysis. In *Proceedings of the 58th annual meeting of the association for computational linguistics*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.293.

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., & Eryiğit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)* (pp. 19–30). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/S16-1002.

Pontiki, M., Galanis, D., Papageorgiou, H., Manandhar, S., & Androutsopoulos, I. (2015). SemEval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015)* (pp. 486–495). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/S15-2082.

Pontiki, M., Galanis, D., Pavlopoulos, J., Papageorgiou, H., Androutsopoulos, I., & Manandhar, S. (2014). SemEval-2014 task 4: Aspect based sentiment analysis. In *Proceedings of the 8th international workshop on semantic evaluation (semeval 2014)* (pp. 27–35). Association for Computational Linguistics, http://dx.doi.org/10.3115/v1/S14-2004.

Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, *20*(1), 61–80. http://dx.doi.org/10.1109/TNN.2008.2005605.

Shi, W., Li, F., Li, J., Fei, H., & Ji, D. (2022). Effective token graph modeling using a novel labeling strategy for structured sentiment analysis. In *Proceedings of the 60th annual meeting of the association for computational linguistics* (pp. 4232–4241). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2022.acl-long.291.

Su, J., Lu, Y., Pan, S., Wen, B., & Liu, Y. (2021). Roformer: Enhanced transformer with rotary position embedding. arXiv:2104.09864, https://api.semanticscholar.org/CorpusID:233307138.

Sun, C., Huang, L., & Qiu, X. (2019). Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 380–385). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/N19-1035.

Tang, F., Fu, L., Yao, B., & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. *Information Sciences*, *488*, 190–204. http://dx.doi.org/10.1016/j.ins.2019.02.064.

Tang, H., Ji, D., Li, C., & Zhou, Q. (2020). Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification. In *Proceedings of the 58th annual meeting of the association for computational linguistics*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.588.

Tang, D., Qin, B., Feng, X., & Liu, T. (2016). Effective LSTMs for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers* (pp. 3298–3307). Osaka, Japan: The COLING 2016 Organizing Committee, https://aclanthology.org/C16-1311.

Varia, S., Wang, S., Halder, K., Vacareanu, R., Ballesteros, M., Benajiba, Y., Anna John, N., Anubhai, R., Muresan, S., & Roth, D. (2023). Instruction tuning for few-shot aspect-based sentiment analysis. In *Proceedings of the 13th workshop on computational approaches to subjectivity, sentiment, & social media analysis* (pp. 19–27). Toronto, Canada: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2023.wassa-1.3.

Wang, M., Cao, D., Li, L., Li, S., & Ji, R. (2014). Microblog sentiment analysis based on cross-media bag-of-words model. In *International conference on internet multimedia computing and service*. https://api.semanticscholar.org/CorpusID:9512214.

Wang, K., Shen, W., Yang, Y., Quan, X., & Wang, R. (2020). Relational graph attention network for aspect-based sentiment analysis. http://dx.doi.org/10.48550/arXiv.2004.12362, arXiv preprint arXiv:2004.12362.

Wang, Y., Zhang, J., Ma, J., Wang, S., & Xiao, J. (2020). Contextualized emotion recognition in conversation as sequence tagging. In *Proceedings of the 21th annual meeting of the special interest group on discourse and dialogue* (pp. 186–195). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.sigdial-1.23.

Wu, S., Fei, H., Ren, Y., Ji, D., & Li, J. (2021). Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge. In *Proceedings of the thirtieth international joint conference on artificial intelligence, IJCAI-21* (pp. 3957–3963). International Joint Conferences on Artificial Intelligence Organization, http://dx.doi.org/10.24963/ijcai.2021/545.

Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., & Xia, R. (2020). Grid tagging scheme for aspect-oriented fine-grained opinion extraction. In *Findings of the association for computational linguistics: EMNLP 2020*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.findings-emnlp.234.

Xiao, Z., Wu, J., Chen, Q., & Deng, C. (2021). BERT4GCN: Using BERT intermediate layers to augment GCN for aspect-based sentiment classification. In *Proceedings of the 2021 conference on empirical methods in natural language processing*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.724.

Xiao, L., Wu, X., Xu, J., Li, W., Jin, C., & He, L. (2024). Atlantis: Aesthetic-oriented multiple granularities fusion network for joint multimodal aspect-based sentiment analysis. *Information Fusion*, *106*, Article 102304. http://dx.doi.org/10.1016/j.inffus.2024.102304.

Xiao, L., Wu, X., Yang, S., Xu, J., Zhou, J., & He, L. (2023). Cross-modal fine-grained alignment and fusion network for multimodal aspect-based sentiment analysis. *Information Processing & Management*, *60*(6), Article 103508. http://dx.doi.org/10.1016/j.ipm.2023.103508.

Xiao, L., Xue, Y., Wang, H., Hu, X., Gu, D., & Zhu, Y. (2022). Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. *Neurocomputing*, *471*, 48–59. http://dx.doi.org/10.1016/j.neucom.2021.10.091.

Xu, L., Chia, Y. K., & Bing, L. (2021). Learning span-level interactions for aspect sentiment triplet extraction. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing* (pp. 4755–4766). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.367.

Yan, H., Dai, J., Ji, T., Qiu, X., & Zhang, Z. (2021). A unified generative framework for aspect-based sentiment analysis. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing* (pp. 2416–2429). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.188.

Ye, S., Zhai, Z., & Li, R. (2023). Enhanced machine reading comprehension method for aspect sentiment quadruplet extraction. In *ECAI 2023* (pp. 2874–2881). IOS Press, http://dx.doi.org/10.3233/FAIA230600.

Zhang, D., Chen, F., & Chen, X. (2023). DualGATs: Dual graph attention networks for emotion recognition in conversations. In *Proceedings of the 61st annual meeting of the association for computational linguistics*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2023.acl-long.408.

Zhang, W., Deng, Y., Li, X., Yuan, Y., Bing, L., & Lam, W. (2021). Aspect sentiment quad prediction as paraphrase generation. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 9209–9219). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.726.

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2021). Towards generative aspect-based sentiment analysis. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing* (pp. 504–510). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-short.64.

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2021). Towards generative aspect-based sentiment analysis. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 2: short papers)* (pp. 504–510). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-short.64.

Zhang, B., Xu, D., Zhang, H., & Li, M. (2019). STCS lexicon: Spectral-clustering-based topic-specific Chinese sentiment lexicon construction for social networks. *IEEE Transactions on Computational Social Systems*, *6*(6), 1180–1189. http://dx.doi.org/10.1109/TCSS.2019.2941344.

Zhao, H., Huang, L., Zhang, R., Lu, Q., & Xue, H. (2020). SpanMlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics*. Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.296.

Zhou, Y., Liao, L., Gao, Y., Jie, Z., & Lu, W. (2021). To be closer: Learning to link up aspects with opinions. In *Proceedings of the 2021 conference on empirical methods in natural language processing* (pp. 3899–3909). Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.317.