Applied Intelligence (2025) 55:740
https://doi.org/10.1007/s10489-025-06313-8

Fpa-GCN: enhancing aspect sentiment triplet extraction with feature-rich prediction-aware graph convolutional networks

Haoyu Jiang[1] · Xiaoliang Chen[1,2,4] · Duoqian Miao[2] · Hongyun Zhang[2] · Xiaolin Qin[3] · Xu Gu[1,3] · Peng Lu[4] · Xianyong Li[1]

Accepted: 28 January 2025 / Published online: 17 June 2025

Abstract

Aspect Sentiment Triplet Extraction (ASTE) is a critical yet challenging task within Aspect-Based Sentiment Analysis (ABSA), which is dedicated to discerning aspect terms, their corresponding opinion terms, and sentiment polarities. The applicability of ASTE spans a broad array of sectors, from e-commerce to social media analytics and customer feedback analysis, highlighting its versatility and significance across various industries. However, the prevailing research methodologies frequently fail to capture the slight interplay among the three essential sentiment elements, often resulting in a misalignment that hampers the establishment of accurate contextual relationships. These misalignments often occur as discrepancies in the model's understanding of contextually related content, deviating from expected associations. To address these challenges, this paper introduces the Feature-Rich Prediction-Aware Graph Convolutional Network (Fpa-GCN) model, meticulously designed for the ASTE task. By harnessing the power of BERT for superior semantic feature extraction and utilizing these features through Biaffine attention, the Fpa-GCN model employs a multi-branch graph convolutional network to effectively encapsulate contextual relations. Additionally, the model incorporates Information Fusion and an advanced Gating mechanism, further enhancing triplet extraction precision. Rigorous evaluations on benchmark datasets affirm the Fpa-GCN model's supremacy over existing state-of-the-art models, achieving significant F1 score improvements. Specifically, the model achieves gains of 1.31%, 0.11%, 3.7%, and 2.15% on the Restaurant14, Laptop14, Restaurant15, and Restaurant16 datasets, respectively, in experiment group $\boldsymbol{D_1}$, and 1.66%, 0.08%, 4%, and 3.33% on the same datasets in experiment group $\boldsymbol{D_2}$. These results underscore the Fpa-GCN model's efficacy and its potential to set a new benchmark in the field of ABSA.

Keywords Natural language processing · Aspect-based sentiment analysis · Aspect sentiment triplet extraction · Graph convolutional network · Biaffine attention

1 Introduction

Aspect sentiment triad extraction (ASTE) has emerged as a pivotal and complex sub-task within aspect-based sentiment analysis (ABSA), aiming to distill aspect-opinion-sentiment terminology triples from textual data. Each triple encapsulates an aspect term (aspects), an opinion term (opinions), and their associated sentiment, offering a granular view of sentiments towards specific aspects within a sentence. Aspects refer to words or phrases directly related to the entity under discussion, while opinions express subjective attitudes. The ASTE task can be practically illustrated using the example of a sentence: "The food is good but the service is bad." shown in Fig. 1. In this case, we can perform a dissection to identify the aspects highlighted in red, the opinions highlighted in green, and their associated sentiment polarities in yellow. The figure exemplifies how a single sentence can hold varied sentiments toward different aspects: while the "food" is regarded positively, the "service" is viewed negatively. The challenge in ASTE lies in the simultaneous extraction of these three interrelated elements. Traditional approaches

✉ Xiaoliang Chen
chenxl@mail.xhu.edu.cn

1 School of Computer and Software Engineering, Xihua University, Chengdu 610039, P. R. China

2 College of Electronic and Information Engineering, Tongji University, Shanghai 201804, P. R. China

3 Chengdu Institute of Computer Applications, Chinese Academy of Sciences, Chengdu 610041, P. R. China

4 Department of Computer Science and Operations Research, University of Montreal, Montreal, QC H3C3J7, Canada

[1, 2] often segment the task into two phases, adopting a pipelined architecture with a unified labeling scheme. This sequential process, however, struggles to capture the intricate relationships among the three elements effectively, leading to potential error propagation. In response, some have proposed end-to-end models for more cohesive triple extraction [3, 4]. Nonetheless, these models primarily focus on developing novel markup strategies that treat ASTE as a singular task without adequately addressing the interconnections between the constituent words. Moreover, the inherent relationships among triples, such as syntactic dependencies and similarities in part-of-speech combinations, are frequently neglected. Our analysis identifies two main shortcomings in existing ASTE models: the underutilization of comprehensive linguistic features and the insufficient exploration of inter-word relationships to improve task performance.

To address these challenges, this study introduces a novel architecture, the Feature-riched Prediction-Aware Graph Convolutional Network (Fpa-GCN). Initially, BERT [5] is utilized as an encoder to obtain contextual embeddings ($H_1 = h_1, h_2, \ldots, h_n$) and ($H^{Bert} = h_1, h_2, \ldots, h_n$). By harmonizing ($H_1$) and ($H^{Bert}$), a Biaffine attention module [6] is applied to (H_1), enabling the prediction of word relationship distributions within sentences through a vector representation, where each dimension signifies a specific relation type. This process yields a relational neighborhood tensor, paving the way for the extraction of lexical combinations, syntactic dependency types, tree-based distances, and relative positional distances for each word pair in every sentence. To manage the noise from embedding these features into a graph, we apply a linear transformation for dimension alignment and introduce a gating mechanism to filter relevant features. Our Fpa-GCN model then constructs a graph from the sentence, treating words and their relational tensors as nodes and edges, respectively, and enriches this graph with multichannel linguistic features. Furthermore, we propose a prediction-aware information fusion (PAIF) module to increase the model's predictive sensitivity. The PAIF module integrates sentence representations (H^{Bert}) from BERT embeddings with augmented linguistic features, channeling them into a layer equipped with a predictive-aware component. This integration of global information with linguistic features through feature-rich graph convolutional operations can enhance the model's predictive accuracy. The mapping prioritizes the learning of task-critical features, thereby elevating overall performance. Our contributions are succinctly summarized as follows:

- **Advancing ASTE with Fpa-GCN.** Introducing the F eature-Enriched Prediction-Aware Graph Convolutional Network (Fpa-GCN), our modular approach integrates BERT and Biaffine attention to intricately model word-context relationships. This combination enhances task performance by employing advanced semantic encoding techniques.
- **Graph-Based Linguistic Feature Extraction.** Our approach advances sentiment analysis by embedding linguistic features into a graph-based structure and applying graph convolution for feature extraction. To mitigate noise and improve feature relevance, we incorporate dimensionality transformation alongside an advanced gating mechanism to further filter essential features. This dual approach addresses the common challenge of noise in data and enhances the relevance of the features extracted.
- **PAIF Module: Enhancing Predictive Sensitivity.** The introduction of Prediction-Aware Information Fusion (PAIF) module represents a significant advancement in improving model sensitivity to predictive information. By fusing BERT-derived sentence representations with enriched linguistic features in a prediction-aware layer, the PAIF module significantly enhances the model's ability to assimilate task-specific features and achieve accurate predictions, focusing on the assimilation of crucial task-specific features.
- **Empirical Validation.** Our comprehensive evaluation on benchmark datasets provides strong evidence that the Fpa-GCN model outperforms existing state-of-the-art baseline models. Specifically, the model achieves notable improvements in F1 scores across several datasets. In experiment group D_1, it demonstrates increases of 1.31%, 0.11%, 3.7%, and 2.15% for the Restaurant14, Laptop14, Restaurant15, and Restaurant16 datasets, respectively. Similarly, in experiment group D_2, the F1 score improvements are 1.66%, 0.08%, 4%, and 3.33% for the same datasets. These results substantiate the superior performance of our method and underscore its robustness across different contexts.

The structure of this paper is organized as follows:

- Section 2 provides an overview of related work, discussing sub-tasks in ABSA and emphasizing the application of GCNs to the ASTE task.
- Section 3 details the proposed Fpa-GCN architecture, explaining its modular design, foundational principles, algorithms, and mathematical formulations.
- Section 4 describes the experimental settings, including datasets, baseline models, primary evaluation metrics, and key experimental results.
- Section 5 presents an in-depth analysis of the model, divided into three subsections: (5.1) Ablation studies assessing the contributions of individual components; (5.2) Case studies illustrating three representative examples; and (5.3) Further analysis examining the model's strengths and limitations from multiple perspectives.

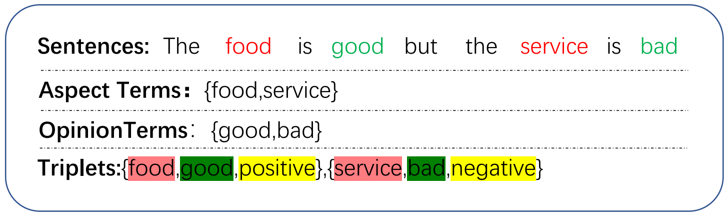


Fig. 1 A case of ASTE. Highlighting aspect terms in red, opinion terms in green, and sentiment polarity in yellow

- Section 6 concludes the paper by summarizing the findings and suggesting directions for future research.

2 Related work

The field of Natural Language Processing (NLP) has seen a growing focus on fine-grained sentiment analysis and opinion mining over the past decade, generating significant research interest. In this section, we will provide a brief review of the evolution of Aspect-Based Sentiment Analysis (ABSA), highlighting the key subtasks within ABSA, including the distinction between single-output and compound-output tasks. Additionally, we will present an overview of the methodologies employed in Aspect Sentiment Triplet Extraction (ASTE) and conclude with insights into the application of Graph Convolutional Networks (GCNs) in ASTE.

2.1 Aspect-based sentiment analysis

Aspect-Based Sentiment Analysis (ABSA) aims to identify both explicit and implicit sentiments within sentences, a task that has seen significant advancements [7–9]. Sentences often contain multiple aspect and opinion terms, reflecting diverse sentiment expressions, particularly prevalent in product and service reviews on e-commerce platforms [10, 11]. Extracting opinions from such reviews is essential for merchants to understand authentic customer feedback, driving substantial research efforts in this area.

ABSA methodologies have evolved from lexicon-based [12], through machine learning [13], to deep learning approaches [14]. Earlier strategies required manual feature engineering, such as bag-of-words [15] and part-of-speech tagging [16], which, while effective, required significant human effort. The emergence of deep learning has revolutionized contextual representation, significantly advancing ABSA tasks [17, 18]. These techniques often adapt pre-trained language models (PLMs) like BERT for task-specific applications, leveraging their superior contextual modeling capabilities. In particular, BERT has had a profound impact on NLP research and serves as the foundation for our proposed model in ABSA.

2.2 ABSA subtasks

This subsection systematically classifies ABSA tasks into two main categories: single -output subtasks and compound -output subtasks, as illustrated in Fig. 2, which provides a clear example to highlight the distinctions between these subtasks. The figure presents the sentence "The drinks are always well made, and the wine selection is fairly priced" as input, identifying two aspect terms-"drinks" (a_1) and "wine selection" (a_2)-along with their corresponding opinion terms-"well made" (O_1) and "fairly priced" (O_2). The sentiment polarities associated with these aspects are both labeled as positive (s_1 and s_2). Following the categorization of ABSA into single -output and compound -output subtasks, we review existing research on each type.

2.2.1 Single output subtasks

A significant portion of ABSA research has focused on single-output subtasks, which aim to extract a specific element—either an aspect (a), sentiment (s), or opinion (o). These subtasks, including Aspect Extraction (AE), Opinion Extraction (OE), Aspect-Level Sentiment Classification (ALSC), and Aspect-Opinion Extraction (AOE), each play a distinct role in advancing the understanding of sentiment analysis. Figure 2 effectively illustrates these four subtasks. AE and OE are concerned with extracting aspect and opinion terms (a_1, a_2 and o_1, o_2) from the input sentence. ALSC classifies the sentiment associated with each aspect, using the aspect term and sentence as input ($s+a_1 \rightarrow s_1, s+a_2 \rightarrow s_2$). AOE extracts the opinion term corresponding to a given aspect ($s + a_1 \rightarrow o_1, s + a_2 \rightarrow o_2$).

Aspect extraction (AE) has consistently been at the forefront of research within the field of Aspect-Based Sentiment Analysis, captivating the interest of numerous scholars. This pivotal task, which aims to identify explicit or implicit aspects from textual data, has traditionally been approached as a sequence tagging challenge [19–21]. The evolution of methodologies in AE, from conventional sequence tagging to the integration of advanced neural network architectures and sequence-to-sequence learning models, reflects the dynamic nature of research in NLP. Among the contributions to this domain, the work by Xu et al. [20] introduced a Convolu-

Categories	Subtasks	Input	Output	Task Type
Single output	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 + a_2	s_1 s_2	Classification
	Aspect-oriented Opinion Extraction(AOE)	S + a_1 S + a_2	o_1 o_2	Extraction
Compound output	Aspect Term Extraction and Sentiment Classification(AESC)	S	$(a_1 + s_1)$ $(a_2 + s_2)$	Extraction &Classification
	Pair Extraction(Pair)	S	$(a_1 + o_1)$ $(a_2 + o_2)$	Extraction
	Aspect Sentiment Triplet Extraction (ASTE)	S	(a_1, o_1, s_1) (a_2, o_2, s_2)	Extraction &Classification

Fig. 2 Illustration of seven ABSA subtasks

tional Neural Network (CNN) model that employs double embeddings for AE, marking a departure from traditional single embedding approaches. This innovative model leverages both word-level and character-level embeddings, providing a richer representation of textual data that enhances the model's ability to capture the nuances of aspect terms. Building on the foundation laid by earlier studies, the field witnessed further advancements with the adoption of sequence-to-sequence learning models, particularly with the integration of pre-trained language models. Ma et al. [22] made a contribution by proposing methodologies that incorporate gated unit networks and a position-aware attention mechanism. The inclusion of gated unit networks and attention mechanisms by Ma et al. has been instrumental in advancing the formalization of AE as a sequence-to-sequence learning task, demonstrating the potential of these models in capturing complex linguistic patterns. Complementing these advancements, Li et al. [23] introduced a masked sequence-to-sequence method that further refines the AE process. Their model, designed to conditionally generate aspect terms based on the input text, represents a sophisticated approach to AE. These studies underscore the vibrant and evolving landscape of research in AE.

Opinion extraction (OE) has traditionally been regarded as a crucial, albeit supplementary, task in sentiment analysis [24–27], essential for gaining a comprehensive understanding of sentiments expressed in text. This task focuses on identifying and extracting opinion terms that are intrinsically linked to aspects within a given text, and various methodologies have been developed to refine and optimize this process. A notable advancement was introduced by Wang et al. [28], who proposed a joint model combining the strengths of recursive neural networks and conditional random fields. This approach deviated from traditional methods by offering a more nuanced, interconnected framework for OE. Expanding on integrated approaches, the Relation-Aware Collaborative Learning (RACL) framework by Chen et al. [26] and the Interactive Multi-task Learning Network (IMN) by He et al. [27] demonstrated the potential of multi-task learning and relation propagation mechanisms to enhance OE. The RACL framework emphasizes fostering a synergistic relationship between aspect and opinion term extraction, representing a significant step toward a more holistic approach to ABSA. Likewise, the IMN model introduces an interactive learning paradigm that simultaneously addresses multiple related tasks, enabling a deeper understanding of the intricate dynamics between aspects and opinions. These methodologies, characterized by the innovative integration of neural network architectures, multi-task learning, and relation-aware mechanisms, have made substantial contributions to the advancement of OE.

Aspect-level sentiment classification (ALSC) represents a nuanced and critical task within the broader field of ABSA, focusing on the identification and categorization of sentiments expressed towards specific aspects in textual data. This task requires a deep understanding of both the contextual relevance of aspects and the sentiment orientation associated with them. A seminal contribution to the field was made by Tang et al. [29], who pioneered the use of Long Short-Term Memory (LSTM) networks to enhance the interaction between aspect terms and their context. Building upon this foundation, subsequent studies have further refined the modeling of aspect-context relations through the integration of attention mechanisms into LSTM-based models [28, 30–32]. These advancements underscore the effectiveness of attention-enhanced LSTM networks in capturing the subtleties of aspect-level sentiments. Moreover, the exploration of Convolutional Neural Networks (CNNs), gated neural networks, and memory neural networks in the context of ALSC has broadened the methodological spectrum, offering diverse and innovative approaches to tackle the complexities of sentiment classification. Each of these methodologies brings a unique perspective to the modeling of aspect-context relations, contributing to the ongoing evolution of ALSC research.

Aspect-opinion extraction (AOE) focuses on the simultaneous identification of aspect and opinion terms from textual data. The introduction of AOE by Fan et al. [33] marked a significant milestone in sentiment analysis research, leading to numerous studies aimed at refining and optimizing the extraction process. The majority of research in this area has gravitated towards employing sequence tagging methods [34, 35], a testament to the effectiveness of these approaches in handling the intricacies of AOE. Among these, the work of Wu et al. [34] stands out for its innovative Grid Tagging Scheme (GTS). The GTS represents a novel approach to AOE, enabling the model to capture the complex relationships between aspect and opinion terms in a structured manner. The incorporation of syntactic structures into AOE models has further enhanced the task's effectiveness. The study by Amir Pouran Ben Veyseh et al. [35] exemplifies this approach, integrating syntactic analysis to improve the model's ability to discern and extract relevant aspect-opinion pairs. The inclusion of syntactic structures underscores the importance of linguistic analysis in deep learning models for sentiment analysis, providing a deeper understanding of the textual context and its impact on sentiment expression.

The comprehensive analysis of single output subtasks described above has established a strong foundation and delivered substantial contributions to the subsequent research on compound output subtasks.

2.2.2 Compound output subtasks

The exploration of compound output subtasks in ABSA represents a significant advancement towards a more comprehensive understanding of sentiments embedded within textual data. This section examines three key subtasks: AESC, Pair, and ASTE. These subtasks are illustrated in Fig. 2. AESC jointly extracts aspects and classifies sentiments, denoted as $((a_1 + s_1), (a_2 + s_2))$. Pair Extraction identifies aspect-opinion pairs, represented as $((a_1 + o_1), (a_2 + o_2))$. Finally, ASTE captures the complete aspect-opinion-sentiment triplets, denoted as $((a_1, o_1, s_1), (a_2, o_2, s_2))$. Researchers have employed various techniques, including unified tagging schemes and multi-task learning frameworks, to push the boundaries of accuracy and efficiency in aspect and sentiment analysis. These compound subtasks reflect a paradigm shift towards more integrated approaches, highlighting the interconnected nature of aspects and sentiments within textual contexts.

Aspect term extraction and sentiment classification (AESC) has garnered significant attention. Researchers have explored various methodologies to tackle the challenges inherent in AESC, ranging from pipeline approaches to innovative unified tagging schemas [36–38] and multi-task learning strategies [26, 27]. These efforts aim to enhance the accuracy of sentiment analysis by addressing error propagation issues and improving the integration of aspect term extraction with sentiment classification. The pipeline method, traditionally employed in AESC tasks, has been complemented and, in some instances, superseded by more integrated approaches. Zhang et al. [37] made a notable contribution to this domain by empirically studying the impact of word embeddings and automatic feature combinations. Their work extends a Conditional Random Field (CRF) baseline through the incorporation of neural networks, showcasing the potential of neural architectures in sentiment analysis. This study marks a significant step towards understanding the intricate dynamics between feature representation and sentiment analysis performance. Li et al. [38] introduced a groundbreaking unified model that tackles target-based sentiment analysis in an end-to-end manner. By employing a unified tagging scheme, their model simplifies the task architecture, reducing the complexity traditionally associated with pipeline methods. In a similar vein, Chen et al. [26] proposed the Relation-Aware Collaborative Learning (RACL) framework. This innovative framework utilizes multi-task learning and relation propagation mechanisms within a stacked multi-layer network, enabling seamless coordination across subtasks. The RACL framework demonstrates the potential of multi-task learning in fostering a synergistic relationship between aspect term extraction and sentiment classification, thereby enhancing the overall effectiveness of the AESC task. He et al. [27] made further advancements with their Interactive Multi-task Learning Network (IMN), which jointly learns multiple related tasks at both the token and document levels. Recent developments also include the introduction of span-based AESC methods [39], which address sentiment inconsistency issues commonly associated with unified tagging schemas. These studies highlight the dynamic and evolving nature of research in AESC. By adopting unified tagging schemes, multi-task learning, and innovative neural network architectures, significant strides have been made in improving the accuracy and efficiency of aspect term extraction and sentiment classification.

Pair extraction (Pair) focuses on identifying (aspect, opinion) pairs within text, and has emerged as a critical area of research. Recent advancements in this area have leveraged sophisticated methodologies, including span-based extraction methods and multi-task learning frameworks, to enhance the accuracy and efficiency of pair extraction. Zhao et al. [40] have made a significant contribution to this field with their proposal of a multi-task learning framework that employs span-based extraction methods for extracting all (aspect, opinion) pair-wise relations from scratch. The methodology introduced by Zhao et al. is commendable for its innovative use of span-based techniques, which have proven to be effective in capturing the intricate relationships between aspects and opinions. Building on the foundation laid by span-based and multi-task learning approaches, Chen et al. [41] introduced the Synchronous Double-channel Recurrent Network (SDRN), a novel architecture designed to enhance the efficiency of pair extraction. The SDRN is complemented by two key mechanisms: the entity synchronization mechanism (ESM) and the relation synchronization mechanism (RSM). These mechanisms are ingeniously designed to extract opinion entities and their relations simultaneously, thereby streamlining the pair extraction process. The methodologies employed in these studies for investigating the pair extraction task have significantly shaped the trajectory of Aspect Sentiment Triplet Extraction (ASTE) research.

Aspect sentiment triplet extraction (ASTE) task, also known as triplet extraction, aims to simultaneously identify aspect terms, opinion terms, and their corresponding sentiment polarities within sentences, presenting a comprehensive approach to understanding sentiments expressed in text. The evolution of methodologies in ASTE reflects the ongoing efforts of NLP researchers to address the complexities of sentiment analysis in a nuanced and integrated manner. Peng et al. [1] were among the pioneers to formally introduce the ASTE task, proposing a two-stage pipeline framework that integrates aspect extraction, aspect sentiment classification, and opinion extraction. Their work laid the groundwork for subsequent research in this area, highlighting the potential of a structured approach to tackling the multifaceted challenges of triplet extraction. Building on the foundational concepts introduced by Peng et al., Mao et al. [2] took a novel approach by reformulating ASTE as a machine reading comprehension problem. By leveraging a shared BERT encoder, their model obtains triplets through multiple stages of decoding, showcasing the adaptability of machine reading comprehension techniques in extracting sentiment-related information from text. Further advancing the exploration of end-to-end methodologies, Xu et al. [42] introduced a novel tagging scheme that capitalizes on explicit local context information to extract triplets in a seamless manner. Their study underscores the critical role of context in determining the performance of ASTE models, emphasizing the need for models to effectively integrate contextual cues for accurate triplet extraction. Complementing these approaches, Wu et al. [34] proposed a grid tagging scheme, akin to table filling, to facilitate the end-to-end extraction of aspect sentiment triplets. This innovative approach addresses the ASTE task by conceptualizing it as a grid-based problem.

The methodologies highlighted above collectively underscore the significant contributions of end-to-end approaches to the advancement of the ASTE task. By emphasizing the effective utilization of relationships among words within sentences, these studies pave the way for more sophisticated and integrated sentiment analysis models. However, it is also noted that these approaches often overlook the intricate connections between words and linguistic features, which could further enhance the effectiveness of triplet extraction. As such, there remains an imperative need for continued research focused on exploring and leveraging these linguistic relationships to optimize the performance of ASTE models.

2.3 Utilization of the GCN in ASTE

The utilization of GCNs has emerged as a critical innovation, particularly in the subtask of ASTE. The syntax dependency tree [43, 44], a critical component in parsing review sentences, is vital in refining key features essential for sentiment analysis. Recent studies have effectively used GCNs to navigate and interpret syntax dependency trees. The conventional application of GCNs in addressing dependency graphs has been well-documented in prior research. Within the scope of ASTE, GCNs have been extensively adopted to amalgamate various sources of information, thereby enriching the analysis with a more nuanced understanding of the syntactic and semantic interplay. Shi et al. [45] applied a GCN to bolster the interaction between syntactic and semantic features, illustrating the network's capacity to bridge these two critical dimensions of language analysis. Further advancing the integration of semantic and syntactic representations, another study [46] employed a GCN module to merge these two facets effectively. This innovative module not only preserved sequential information but also amplified the linguistic representation, showcasing the versatility of GCNs in enhancing the depth and accuracy of sentiment analysis.

Addressing the challenge of multiple aspect terms corresponding to a single opinion term (and vice versa), Li et al. [47] integrated a GCN with a base encoder to construct comprehensive span representations. This approach adeptly covers both aspect and opinion terms, demonstrating the GCN's utility in capturing the multifaceted relationships inherent in sentiment expressions. Similarly, Fei et al. [48]

utilized a GCN to model the graph based on concatenated representations of aspect and opinion terms, further highlighting the indispensable role of GCNs in refining feature representations within ASTE.

Previous applications of GCN in ASTE have achieved significant progress but face notable limitations. A primary challenge is their heavy reliance on syntactic dependency trees. While these structures are useful for modeling relationships between aspect and opinion terms, their effectiveness depends on the accuracy of the parsers that generate them. In informal, ungrammatical, or domain-specific texts-such as user-generated reviews containing slang or typographical errors-parsing inaccuracies can propagate through the model, resulting in suboptimal performance. For instance, incorrect dependency relations may impede the accurate association of aspect and opinion terms. Another critical limitation is the inadequate incorporation of semantic features. Many existing approaches prioritize syntactic information from dependency graphs, often overlooking semantic nuances essential for capturing implicit opinions and contextual sentiment. For example, a sentence like “The battery lasts all day, a rare feat for devices in this price range” requires a deeper semantic understanding that GCNs relying solely on syntactic features may struggle to provide. To address these challenges, our proposed Fpa-GCN introduces a more robust framework for context-aware and semantically enriched ASTE models, effectively mitigating the limitations of earlier methods. Fpa-GCN is a novel model that draws inspiration from the successful employment of BERT and GCNs in NLP tasks. This model aims to harness the synergistic potential of BERT's deep contextual insights and GCNs' adept handling of dependency structures, setting a new benchmark in the field of ABSA.

3 Method

This section delineates the methodology underpinning the feature-riched prediction-aware graph convolutional network (Fpa-GCN) framework, designed to address the ASTE task. Our exposition commences with a precise formulation of the ASTE problem, setting the stage for a deeper exploration of the Fpa-GCN architecture. First, we proceed to elucidate the initial phase of our methodology, focusing on the input and encoding layer where BERT is leveraged to generate foundational semantic features. This is followed by an examination of the Biaffine attention mechanism. This critical component refines semantic features by capturing nuanced relationships between words. Second, we introduce the GCN layer after introducing the attention mechanism. This layer is instrumental in integrating semantic features with four additional linguistic feature representations, enriching the model's understanding of the text. Third, a gating mechanism is employed at this juncture to selectively filter and refine the features, ensuring that only the most relevant information is propagated through the network. The core of our methodology is encapsulated in the Prediction-Aware Information Fusion (PAIF) subsection. Here, we detail how the PAIF module synergizes the refined features with predictive cues, enhancing the model's ability to anticipate sentiment orientations accurately. The overall architecture of Fpa-GCN is depicted in Fig. 3.

3.1 Problem formulation

The ASTE task is formulated to dissect a given sentence into its constituent sentiments. Specifically, a sentence is represented as a sequence of words $X = \{w_1, w_2, \ldots, w_n\}$, where n denotes the total number of words. The core objective of ASTE is to meticulously analyze the input sentence and yield a set of sentiment-oriented triplets $f(a, o, s)^m$, where each triplet consists of an aspect term a, an opinion term o, and the corresponding sentiment polarity s. Here, m represents the aggregate count of triplets extracted from the sentence. The aspect term a is further delineated into the tuple $(B-a, E-a)$, where B and E signify the beginning and ending indices of the aspect within the sentence, respectively. Analogously, the opinion term o is detailed as $(B-o, E-o)$, encapsulating the start and end positions of the opinion. The sentiment polarity s is categorized into one of three predefined classes: positive, neutral, or negative. This classification reflects the sentiment conveyed by the opinion term towards the associated aspect term. For instance, as illustrated in Fig. 1, the extracted triplets include $(food, good, positive)$ and $(service, bad, negative)$, demonstrating the model's capability to discern diverse sentiment orientations within a single sentence. To further elucidate the ASTE task, we introduce a comprehensive set of ten specific relationships among words within a review, as detailed in Table 1. These relationships, designated as labels, signify the intricate connections between pairs of words and embody the final predictions output by our proposed model, the Fine-grained Position-Aware Graph Convolutional Network (Fpa-GCN).

3.2 Input and encoding layer

In the architecture of our Fpa-GCN model, the foundational layer is the input and encoding mechanism, which is powered by BERT [5]. Specifically, we employ the bert-base-uncased variant to distill semantic features from the input text. This encoding layer is tasked with generating a sequence of hidden representations $H = \{h_1, h_2, \ldots, h_n\}$, derived from the

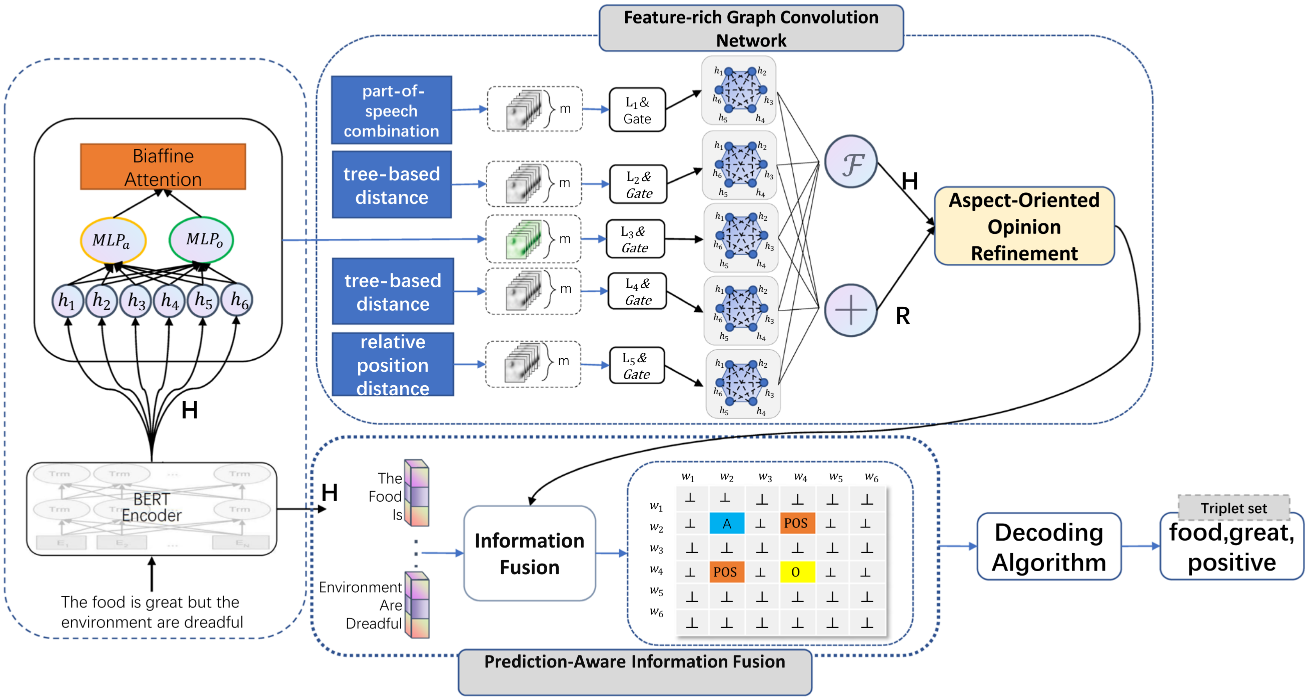


Fig. 3 The overall framework of the proposed feature-riched prediction-aware graph convolutional network

final transformer block, for an input sentence X composed of n tokens $\{w_1, w_2, \ldots, w_n\}$. This process ensures that the contextual nuances of each token within the sentence are captured, laying a robust foundation for subsequent analysis.

3.3 Biaffine attention

To adeptly capture the intricate relationships between word pairs within a sentence, our model incorporates a biaffine attention module. This choice is inspired by the module's proven efficacy in syntactic dependency parsing tasks [6]. The biaffine attention mechanism operates by first extracting the hidden states h_α and h_β of any two words w_α and w_β in the sentence X. These states are retrieved from h_λ, where $\lambda \in [1, 12]$ denotes the layer within the transformer, and h_0 represents the input embedding produced by BERT, which is further refined through a linear function. The outputs across the 12-layer transformer are denoted as $[h_1, h_2, \ldots, h_{12}]$. The biaffine attention process is mathematically articulated as follows:

$$h_\alpha^a = MLP_a(h_i), \tag{1}$$

$$h_\beta^o = MLP_o(h_j), \tag{2}$$

Table 1 Overview of defined relations for ASTE tasks

Items	Segment	Relation	Definition
1	Aspect term detection	B-A	Beginning of an aspect term.
2		E-A	End of an aspect term.
3		A	An aspect term.
4	Opinion term detection	B-O	Beginning of an opinion term.
5		E-O	End of an opinion term.
6		O	An opinion term.
7	Sentiment classification	POS	Sentiment polarity is positive.
8		NEU	Sentiment polarity is neutral.
9		NEG	Sentiment polarity is negative.
10	Default	*	Does not belong to any aforementioned categories.

Through the application of MLP_α and MLP_β, we derive the aspect-specific feature h^a_α (1) and the opinion-specific feature h^o_β (2). These features are then seamlessly integrated into the Biaffine attention mechanism, which undergoes a transformation process defined as:

$$\xi_{\alpha,\beta} = {h^a_\alpha}^\top \mu_1 h^o_\beta + \mu_2(h^a_\alpha \oplus h^o_\beta) + b, \tag{3}$$

$$r_{\alpha,\beta,\gamma} = \frac{\exp(g_{\alpha,\beta,\gamma})}{\sum_{\lambda=1}^{m} \exp(g_{\alpha,\beta,\lambda})}, \tag{4}$$

$$R = Biaffine(MLP_a(H_1), (MLP_o(H_1)), \tag{5}$$

where, μ_1, μ_2, and b represent the trainable weights and biases, and $\oplus$ signifies the concatenation operation. The score vector $r_{\alpha,\beta} \in \mathbb{R}^{1\times m}$ encapsulates the relational dynamics between w_i and w_j, with m indicating the total number of relation types. Furthermore, $r_{\alpha,\beta,\gamma}$ quantifies the score attributed to the γ-th relation type for the word pair (w_α, w_β). The resultant adjacency tensor $R \in \mathbb{R}^{n\times n\times m}$ models the relational interplay between words within the sentence. By leveraging the biaffine attention's dual focus on aspect and opinion features, our model proficiently predicts the relational probability R_{ba} between word pairs. This relational data is subsequently integrated with four additional types of linguistic features through the application of GCNs, enriching the model's analytical capabilities.

3.4 Feature-rich graph convolution network

In the advancement of NLP, particularly in the context of ASTE, text representation transcends mere semantic feature mapping. It encompasses a broader spectrum of linguistic features, among which the syntax dependency graph is essential. This graph, symbolized as $G = (V, E)$, is composed of vertices V (representing the words or nodes) and edges E (denoting the syntactic relations or dependencies between these nodes). These syntactic relationships are typically encoded using an adjacency matrix A, where $A_{\alpha,\beta}$ is assigned a value of 1 to signify a syntactic connection between w_α and w_β, and 0 in its absence. To enrich the contextual representation of sentences, this study introduces three linguistic features for each word pair: part-of-speech combinations, tree-based distances, and relative positional distances, as illustrated in Fig. 4. Prior to their integration into the GCN, these features are encoded as adjacency matrices.

A gating mechanism is subsequently applied to both R_{ba} and the four categories of linguistic features. For example, this mechanism selectively filters the features encoded by the syntactic dependency type tensor E_{sdt}, emphasizing the extraction of more essential features. The process is delineated as follows:

$$R^{sdt}_{i,j} = \{r^{sdt}_{i,j} \mid 1 \le i, j \le n\}, \tag{6}$$

$$r'^{sdt}_{i,j} = LayerNorm(r^{sdt}_{i,j}), \tag{7}$$

$$g^{in}_{i,j}, g^{out}_{i,j} = sigmoid(Linear(r'^{sdt}_{i,j})), \tag{8}$$

$$tempt_{ij} = g^{in}_{i,j} \odot Linear(r'^{sdt}_{i,j}), \tag{9}$$

$$\tilde{r}^{sdt}_{i,j} = tempt_{ij} \odot g^{out}_{i,j}, \tag{10}$$

$$\tilde{R}^{sdt}_{i,j} = \{\tilde{r}^{sdt}_{i,j} \mid 1 \le i, j \le n\}, \tag{11}$$

This mechanism operates by learning a set of gating values that modulate the importance of each feature. Specifically, the syntactic dependency type tensor E_{sdt} undergoes this filtering process to emphasize the extraction of more critical features. First, the syntactic dependency features $r^{sdt}_{i,j}$ are extracted and normalized as shown in (6) and (7), where $r^{sdt}_{i,j}$ represents the raw features and LayerNorm applies normalization for stable training.

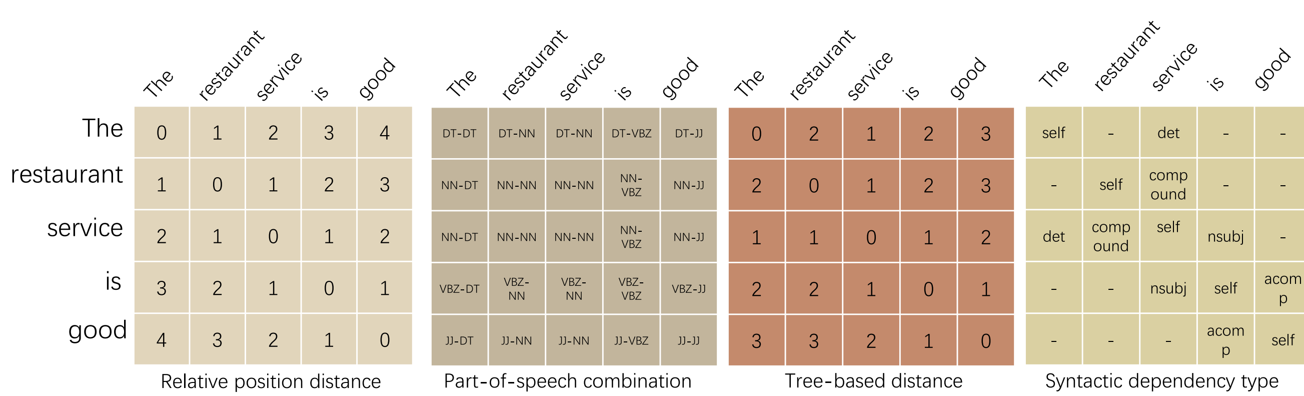


Fig. 4 The example of four mentioned types of dependency relations among words in reviews

Next, the gating mechanism computes the input and output gates, $g_{i,j}^{in}$ and $g_{i,j}^{out}$, using a sigmoid function applied to the normalized tensor $r_{i,j}^{\prime sdt}$, as shown in (8). This allows the mechanism to selectively filter the features by learning which parts of the tensor should be emphasized or suppressed.

The next step involves computing a temporary feature tensor, $tempt_{ij}$, by applying the input gate $g_{i,j}^{in}$ to the linearly transformed features, as outlined in (9). The output gate $g_{i,j}^{out}$ is then used to refine the final filtered features, $\tilde{r}_{i,j}^{sdt}$, as defined in (10).

Finally, the filtered syntactic dependency tensor, $\tilde{R}_{i,j}^{sdt}$, is obtained by collecting the refined features across all pairs of indices (i, j), as shown in (11). This final output serves as the refined representation of the syntactic dependency features, ready to be integrated into subsequent processing steps.

Subsequently, a GCN is employed to systematically aggregate information across the individual channels corresponding to each node. This process is encapsulated as follows:

$$\tilde{H}_k^{sdt} = \sigma(\tilde{R}_k^{sdt} H W_k + b_k), \tag{12}$$

$$\tilde{H}^{sdt} = f(\tilde{H}_1^{sdt}, \tilde{H}_2^{sdt}, \ldots, \tilde{H}_m^{sdt}), \tag{13}$$

where W_k and b_k represent the trainable weight and bias, respectively, and σ denotes the activation function (e.g., ReLU). The node hidden representations are processed through the average pooling function $f(\cdot)$. Similarly, representations for $\tilde{H}^{psc}$, $\tilde{H}^{tbd}$, $\tilde{H}^{rpd}$, and $\tilde{H}^{ba}$ are obtained.

The final step involves employing the average pooling function and concatenation operation on node and edge representations, respectively:

$$H = f(\tilde{H}^{psc}, \tilde{H}^{tbd}, \tilde{H}^{rpd}, \tilde{H}^{ba}, \tilde{H}^{sdt}), \tag{14}$$

$$R = \tilde{R}^{psc} \oplus \tilde{R}^{tbd} \oplus \tilde{R}^{rpd} \oplus \tilde{R}^{ba} \oplus \tilde{R}^{sdt}, \tag{15}$$

Here,$H = \{h_1, h_2, \ldots, h_n\}$ and $R = \{r_1, r_2, \ldots, r_n\}$ symbolize the node and edge representations of word pairs, respectively. For label prediction, the representation of the word pair (w_α, w_β) is obtained by concatenating their respective node representations, h_α and h_β, along with their edge representation, $r_{\alpha\beta}$. To refine the precision of the extracted relations, a refinement strategy is implemented to enhance the connections among words, considering the implicit results of aspect and opinion extraction when assessing the compatibility of word pairs. Thus, the refined representation of the word pair is expressed as:

$$\psi_{ij} = h_\alpha \oplus h_\beta \oplus r_{\alpha\beta} \oplus r_{\alpha\alpha} \oplus r_{\beta\beta}, \tag{16}$$

This comprehensive approach not only encapsulates the multifaceted nature of linguistic feature integration but also underscores the model's capability to discern and refine the intricate relationships within the input text.

3.5 Prediction-aware information fusion

In our model, we employ an advanced technique known as Prediction-Aware Information Fusion. This method synergistically combines attributes derived from BERT embeddings with global information and insights gleaned from a feature-rich GCN. The fusion process is mathematically represented as follows:

$$\zeta_{ij} = \psi \oplus H^{Bert}, \tag{17}$$

where ζ_{ij} denotes the fused representation of a word pair. This representation is subsequently passed through a linear layer, followed by the application of a softmax function to produce a label probability distribution ι_{ij}, encapsulated by the equation:

$$\iota_{ij} = softmax(W_p \zeta_{ij} + b_p), \tag{18}$$

where W_p and b_p signify the learnable weight and bias, respectively. This process ensures a comprehensive integration of diverse information sources, enhancing the model's predictive capabilities.

3.6 Loss function

The design of an effective loss function is crucial for optimizing deep learning models. While cross-entropy is a commonly employed loss function, our proposed Fpa-GCN model necessitates a more nuanced approach. This is due to the incorporation of diverse contextual information, which must be reflected in the fine-tuning process. To address this, we introduce an isolated loss, L_{ba}, specifically designed to evaluate the impact of R_{ba}, as shown below:

$$\mathcal{L}_{ba} = -\sum_{i}^{n}\sum_{j}^{n}\sum_{c\in C}\mathbb{I}(y_{ij} = c)\log(\iota_{i,j\|c}), \tag{19}$$

Here, $\mathbb{I}(\cdot)$ denotes the indicator function, $y_{i,j}$ represents the actual relation for the word pair (w_α, w_β), and C encompasses the complete set of relations. Employing a similar methodology, losses for the four distinct linguistic features, namely $\mathcal{L}_{psc}$, $\mathcal{L}_{sdt}$, $\mathcal{L}_{tbd}$, and $\mathcal{L}_{rpd}$, are computed. The final loss function $\mathcal{L}$, integrating these components, is formulated as:

$$\mathcal{L} = \mathcal{L}_p + \gamma \mathcal{L}_{ba} + \delta(\mathcal{L}_{psc} + \mathcal{L}_{sdt} + \mathcal{L}_{tbd} + \mathcal{L}_{rpd}), \quad (20)$$

where γ and δ serve as coefficients to balance the influence of the respective relation constraint losses. The task of ASTE utilizes the conventional cross-entropy loss, $\mathcal{L}_p$, for the primary prediction task, as defined by:

$$\mathcal{L}_p = -\sum_{i}^{n}\sum_{j}^{n}\sum_{c \in C} \mathbb{I}(y_{ij} = c) \log(p_{i,j\|c}), \quad (21)$$

This comprehensive loss function framework can ensure that the model not only learns from the primary task but also benefits from the nuanced understanding provided by the additional linguistic features, thereby achieving a more accurate prediction capability.

3.7 Model training

In this section, we elucidate the training process of the Fpa-GCN, a sophisticated model designed for the ASTE task. The training algorithm is meticulously crafted to optimize the Fpa-GCN model, ensuring it accurately extracts sentiment-oriented triplets from textual data. The overall procedure is encapsulated in Algorithm 1, which we detail below.

The training commences with the input of a sentence set $\mathcal{S}$ and its corresponding label set $\mathcal{L}_\mathcal{S}$, alongside a set of truth triplets $\mathcal{T}_\mathcal{S}$ derived from the sentence set. The objective is to train the Fpa-GCN model to output triplets $\mathcal{T}$ for a given sentence, thereby refining the model's ability to discern and extract relevant sentiment triplets. The core training loop is iterative until the model converges to an optimal performance state. Within each iteration, the Fpa-GCN model undergoes training with the provided sentence and label sets alongside the truth triplets. This iterative process is succinctly represented as:

1. **Repeat**
2. train(FPAGCN($(\mathcal{S}, \mathcal{L}_\mathcal{S}), \mathcal{T}_\mathcal{S}$))
3. **Until Convergence**

Upon convergence, the model is deemed trained, and the triplets $\mathcal{T}$ are extracted using the Fpa-GCN model with the sentence set $\mathcal{S}$ and the truth triplet set $\mathcal{T}_\mathcal{S}$ as inputs.

The function **FPAGCN** encapsulates the essence of the model's operation. Initially, the function **Encoder** is invoked to generate embeddings H and H^{Bert} from the sentence set $\mathcal{S}$. Following this, linguistic features are extracted using the function **LinguisticFeatures**, yielding representations such

Algorithm 1 Overall procedure of Fpa-GCN.

Input: Sentence set $\mathcal{S}$ and label set $\mathcal{L}_\mathcal{S}$,tset of truth triplets from the sentence set $\mathcal{T}_\mathcal{S}$
Output: Triplets $\mathcal{T}$ of the given sentence and Trained model Fpa-GCN

repeat
    train(FPAGCN($(\mathcal{S}, \mathcal{L}_\mathcal{S}), \mathcal{T}_\mathcal{S}$))
until Convergence
$Triplets \leftarrow FPAGCN(\mathcal{S}, \mathcal{T}_\mathcal{S})$
return Trained model Fpa-GCN,Triplets
function FPAGCN($\mathcal{S}$)
    $H, H^{Bert} \leftarrow$ Encoder($\mathcal{S}$)
    $R^{psc}, R^{rpd}, R^{tbd}, R^{sdt} \leftarrow$ LinguisticFeatures($\mathcal{S}$)
    $R^{ba} \leftarrow$ Biaffine(H, H)
    $R^{all} \leftarrow$ GatingMechanism ($R^{psc}, R^{sdt}, R^{tbd}, R^{rpd}, R^{ba}$)
    $\tilde{H} \leftarrow$ MultiLayerGCN(H, R^{all})
    $R \leftarrow$ Concat($R^{psc}, R^{rpd}, R^{tbd}, R^{sdt}, R^{ba}$)
    $H^{psc}, H^{sdt}, H^{tbd}, H^{rpd}, H^{ba} \leftarrow$ Equations (12) and (13)
    $H \leftarrow$ AveragePooling($H^{psc}, H^{sdt}, H^{tbd}, H^{rpd}, H^{ba}$)
    $\zeta_{ij} \leftarrow$ (16), (17)
    $P \leftarrow$ GetPrediction(ζ_{ij})
    $Triplet \leftarrow$ TripletExtraction(P)
    return $Triplets$
end function
function Encoder($\mathcal{S}, \mathcal{L}_\mathcal{S}$)
    $H \leftarrow$ Bert($\mathcal{S}, \mathcal{L}_\mathcal{S}$)
    $H^{Bert} \leftarrow$ Bert($\mathcal{S}, \mathcal{L}_\mathcal{S}$)
    return H,H^{Bert}
end function
function LinguisticFeatures($\mathcal{S}$)
    $R^{tbd}, R^{sdt} \leftarrow$ Supar($\mathcal{S}$)
    $R^{psc} \leftarrow$ Stanza($\mathcal{S}$)
    $R^{rpd} \leftarrow$ GetRelativeDistance($\mathcal{S}$)
    return $R^{psc}, R^{sdt}, R^{tbd}, R^{rpd}$
end function
function GetPrediction(Z)
    return (18)
end function

as R^{psc}, R^{rpd}, R^{tbd}, and R^{sdt}. These features, along with the Biaffine attention representation R^{ba} obtained from applying the function **Biaffine** to H, are then processed through a **GatingMechanism**. This mechanism selectively filters the features, ensuring that only the most relevant ones are forwarded for further processing.

Subsequently, a function **MultiLayerGCN** applies graph convolution over the features and embeddings, producing an enriched representation $\tilde{H}$. This representation, along with the concatenated features R, undergoes pooling to yield a unified feature representation H. The prediction-aware information fusion is then applied, resulting in a prediction matrix ζ_{ij}, from which the final predictions P are derived using the function **GetPrediction**. Finally, the function **TripletExtraction** decodes these predictions into the desired triplets $\mathcal{T}$.

The **Encoder**, **LinguisticFeatures**, and **GetPrediction** functions are auxiliary components of the Fpa-GCN model, each playing a crucial role in the model's operation. The *Encoder* function utilizes BERT to generate embeddings for the sentence set $\mathcal{S}$ and its labels $\mathcal{L}_\mathcal{S}$. The function **LinguisticFeatures** leverages tools such as Supar and Stanza to extract syntactic and part-of-speech features, while the function *GetPrediction* computes the final prediction scores based on the information fusion results.

This comprehensive training algorithm ensures that the Fpa-GCN model is optimally trained to perform the ASTE task, leveraging both semantic and linguistic features to accurately extract sentiment-oriented triplets from textual data.

3.8 Algorithm complexity analysis of training

The time complexity analysis of the Fpa-GCN involves a detailed examination of each component and operation within the Algorithm 1. This analysis aids in understanding the computational efficiency and scalability of the proposed method.

- Encoding Layer: The encoding layer employs BERT to transform the input sentences $\mathcal{S}$ into hidden representations H and H^{Bert}. The time complexity of BERT depends on the number of tokens in the input sentences, the model architecture (specifically the number of layers L, hidden units H, and self-attention heads A), resulting in a complexity of $O(nLHA)$, where n is the average length of the sentences in the set $\mathcal{S}$.
- Linguistic Features Extraction: The extraction of linguistic features involves processing the input sentence set $\mathcal{S}$ through natural language processing tools such as Supar and Stanza. Assuming that these tools process each sentence linearly with respect to the sentence length, the time complexity for this step can be approximated as $O(n)$ for each sentence. Since these features are extracted once per sentence and are independent of other sentences, the total complexity for this step remains $O(n)$ when averaged over the dataset.
- Biaffine Attention: The biaffine attention module involves pairwise interactions between tokens in a sentence. For a sentence of length n, the module computes interactions between all pairs of tokens, leading to a time complexity of $O(n^2)$. However, since this is applied once per sentence, the overall complexity across all sentences averages out, maintaining $O(n^2)$.
- Gating Mechanism and GCN Integration: The gating mechanism processes each linguistic feature individually, followed by the application of a multi-layer GCN on the combined features. If we denote d as the dimensionality of the features and k as the number of GCN layers, the complexity of processing each linguistic feature is $O(nd)$. The subsequent GCN integration has a complexity of $O(kd^2)$ for each layer, leading to a total complexity of $O(knd^2)$ for all features across all sentences.
- Information Fusion and Prediction: The information fusion step involves concatenating feature representations and passing them through a prediction layer, which generally operates in linear time with respect to the number of features and the length of the sentences. Thus, this step has a complexity of $O(nd)$, where d is the dimensionality of the fused feature set.
- Triplet Extraction: The final step, triplet extraction, typically involves parsing through the predictions for each token pair in the sentence to extract the aspect, opinion, and sentiment triplets. This step has a time complexity of $O(n^2)$, considering all possible pairs in sentences of length n.

Summing up all components, the overall time complexity of the Fpa-GCN training algorithm can be approximated as $O(nLHA) + O(n) + O(n^2) + O(knd^2) + O(nd) + O(n^2)$ per sentence. Since $O(n^2)$ and $O(knd^2)$ terms are likely to dominate for longer sentences and larger feature dimensions, the time complexity can be further approximated as $O(n^2) + O(knd^2)$ per sentence, with actual costs depending on the specific dimensions of the input and model parameters.

It is worth noting that while the time complexity per sentence is quadratic with respect to the sentence length, in practice, the lengths of sentences are typically bounded, and the implementation efficiencies, especially those related to matrix operations in BERT and GCN, can significantly mitigate computational costs. However, the scalability of Fpa-GCN may still be challenged by very large datasets or exceedingly long sentences, thus necessitating efficient batching, parallelization, and hardware acceleration techniques for practical deployment.

4 Experiment settings

This section presents the outcomes of the conducted experiments through tables and figures, accompanied by detailed analyses. Initially, we introduce the ASTE datasets widely employed and the associated experimental configurations. Subsequently, a thorough presentation of the experimental results is provided.

4.1 Datasets

This study conducts extensive experiments on four publicly available benchmarks for ASTE: Laptop14, Restaurant14, Restaurant15, and Restaurant16. These datasets originate from the SemEval series of ABSA challenges, specifically the editions held in 2014, 2015, and 2016 [49–51]. The significance of these datasets lies in their comprehensive coverage of consumer reviews, focusing on two main domains: laptops and restaurants. This diversity allows for a robust evaluation of ASTE models across different contexts and terminologies.

For the purpose of ASTE tasks, these datasets have been revised by [34] and [42], hereafter referred to as $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. These revisions were undertaken to adapt the original ABSA datasets for the specific requirements of ASTE, including the identification and extraction of aspect terms, sentiment polarities, and the relationships between them. The modifications introduced in these versions facilitate a more precise and task-oriented evaluation of ASTE models.

Table 2 presents the statistics for these two groups of experiment datasets, offering a detailed breakdown of the number of sentences (#S) and the number of triplets (#T) contained within each dataset. The datasets are further divided into training, development (dev), and test sets, ensuring a comprehensive evaluation framework that supports both model training and performance assessment.

For $\mathcal{D}_1$, the Laptop14 dataset comprises 899 training sentences with 1452 triplets, 225 development sentences with 383 triplets, and 332 test sentences with 547 triplets. The Restaurant14 dataset includes 1259 training sentences with 2356 triplets, 315 development sentences with 580 triplets, and 493 test sentences with 1008 triplets. The Restaurant15 and Restaurant16 datasets follow a similar structure, with detailed statistics provided in Table 2.

The $\mathcal{D}_2$ version of the datasets exhibit slight variations in the number of sentences and triplets, reflecting the different revisions made by [42]. For instance, the Laptop14 dataset in $\mathcal{D}_2$ contains 906 training sentences with 1460 triplets, indicating a minor increase compared to $\mathcal{D}_1$.

4.2 Baselines

Our proposed model Fpa-GCN is benchmarked against a diverse array of state-of-the-art (SOTA) models, each designed with unique approaches to tackle the ASTE task. Below, we provide a comprehensive overview of these baseline models, highlighting their methodologies and contributions to the domain of ASTE.

- GTS-BERT [34] integrates the Grid Tagging Scheme (GTS) with BERT, utilizing BERT's powerful contextual embeddings to enhance aspect-sentiment pair extraction. This model significantly improves the efficiency and accuracy of aspect-sentiment identification, setting a high benchmark for ASTE tasks by effectively leveraging pre-trained contextual embeddings to capture fine-grained relationships between aspects and sentiments.
- GTS-CNN [34] Similar to GTS-BERT, GTS-CNN employs the Grid Tagging Scheme but replaces BERT with a Convolutional Neural Network (CNN) for feature extraction. This approach highlights the efficacy of CNNs in capturing local dependencies within the text, offering a complementary method to aspect-sentiment extraction. By incorporating the grid-based tagging mechanism, GTS-CNN further enhances the robustness of sentiment extraction, demonstrating the power of CNNs in capturing spatial dependencies.
- GTS-BiLSTM [34] combines the Grid Tagging Scheme with a Bidirectional Long Short-Term Memory (BiLSTM) network, improving the model's ability to capture both forward and backward contextual information. This enables the model to better handle complex sentence structures and sequential dependencies, making it particularly effective for modeling syntactic nuances in ASTE tasks.
- S^3E^2 [46] model introduces a graph-sequence dual representation to capture both syntactic and semantic relationships within sentences. By combining graph and sequence representations, this model offers a more nuanced method for understanding sentence dynamics, pushing the boundaries of traditional models and enhancing the ability to model complex sentence structures.
- Peng-two-stage+IOG [34] integrates the Peng-two-stage pipeline with the Incremental Opinion Graph (IOG), providing an innovative approach to aspect-sentiment triplet extraction. By combining structured extraction methods with graph-based models, the Peng-two-stage+IOG model improves the assembly of triplets, offering a highly structured and effective pipeline for aspect-sentiment extraction.
- Peng-two-stage [1] employs a sequential approach to aspect-sentiment extraction, with a clear division of

Table 2 Comparative statistics of datasets across two experimental groups

Datasets		Laptop14		Restaurant14		Restaurant15		Restaurant16	
		#S	#T	#S	#T	#S	#T	#S	#T
$\mathcal{D}_1$	train	899	1452	1259	2356	603	1038	863	1421
	dev	225	383	315	580	151	239	216	348
	test	332	547	493	1008	325	493	328	525
$\mathcal{D}_2$	train	906	1460	1266	2338	605	1013	857	1394
	dev	219	346	310	577	148	249	210	339
	test	328	543	492	994	322	485	326	514

[1]Note: #S denotes the number of sentences; #T means the number of triplets contained in the datasets

tasks between the two stages. This approach effectively addresses the complexity of aspect-sentiment triplet extraction, providing a strong foundation for understanding intricate sentiment structures in text.

- OTE-MTL [52] utilizes a multi-task learning framework to simultaneously extract aspect terms, opinion terms, and their corresponding sentiments. This joint optimization improves the model's ability to learn interdependencies across different aspects of sentiment analysis, enhancing overall performance by learning shared representations across tasks.
- JET-BERT [42] introduces a position-aware tagging mechanism, allowing for the simultaneous extraction of aspect terms, opinion terms, and sentiments. This innovative tagging approach improves the precision and recall of sentiment extraction, particularly in managing the positional relationships between aspects and sentiments, and represents a significant advancement in the ability to capture complex relationships in text.
- BMRC [2] redefines the ASTE task as a Multi-Turn Machine Reading Comprehension (MTMRC) challenge, transforming the process of extracting aspect-sentiment triplets into a dynamic, interactive process. This novel framework broadens the scope of ASTE by framing it within an interactive question-answering paradigm, offering a new way to approach sentiment extraction.
- EMC-GCN [53] represents sentences as multi-channel graphs, where words and relations are treated as nodes and edges. This graph-based approach allows for a rich representation of sentence structure, effectively modeling complex dependencies between aspect-sentiment pairs. Notably, EMC-GCN leverages graph convolution to capture intricate relationships, aligning closely with Fpa-GCN, and highlighting the power of graph-based representations in sentiment analysis.
- MuG-Bert [54] combines multi-task learning with a grid decoding mechanism, inspired by the GTS framework. This model showcases how the combination of multi-task learning and structured decoding can improve aspect-sentiment triplet extraction, building on the strengths of previous models like GTS-BERT to enhance extraction precision.
- UniASTEBERT [55] decomposes the ASTE task into three subtasks: target tagging, opinion tagging, and sentiment tagging. This approach allows for more focused sentiment extraction, improving accuracy and granularity by refining the process of handling complex aspect-sentiment relationships.
- Dual-MRC [56] constructs two machine reading comprehension tasks, training two BERT-based models with shared parameters. This dual-perspective approach exemplifies the power of multi-perspective learning, enhancing the model's ability to extract aspect-sentiment triplets from multiple viewpoints, thereby enriching the overall extraction process.

These baseline models represent the cutting edge of research in ASTE, each contributing invaluable insights and methodologies to the field. Many of these models, such as EMC-GCN and GTS-BERT, share structural similarities with Fpa-GCN, particularly in their use of graph convolution or grid-based mechanisms. The comparison not only highlights the effectiveness of Fpa-GCN but also positions it as part of an ongoing evolution in ASTE, building upon and advancing the approaches established by these SOTA models.

4.3 Implementation details

This section delineates the technical specifications and configurations employed during the experimental phase of our study.

1. Hyperparameter Configuration
   The model training process is governed by the following key hyperparameters, which are selected to ensure optimal performance while balancing computational efficiency and model generalization:
   - Batch Size: The batch size is set to 16. This value is chosen to balance the trade-off between the training speed and memory consumption during training, considering the computational limitations of typical hardware setups.
   - Number of Epochs: The model undergoes training for a total of 100 epochs. This duration ensures the model receives sufficient exposure to the data while preventing underfitting.
   - Learning Rate: The learning rate for updating all trainable parameters is set to 10^{-3}. This choice allows for a sufficiently fast convergence while minimizing the risk of instability during optimization. For the BERT component, a lower learning rate is employed to prevent disrupting the pre-trained weights.
   - BERT Fine-tuning Learning Rate: The learning rate for fine-tuning the BERT encoder is set to 2×10^{-5}. This value is selected to ensure that the pre-trained knowledge embedded in the BERT model is preserved, allowing for gradual adaptation to the specific task.
   - Adam Optimizer Epsilon: The epsilon parameter of the Adam optimizer is set to 1×10^{-8} to enhance numerical stability and prevent issues arising from very small gradient values during optimization.
   - Weight Decay: The weight decay coefficient is set to 0, as no explicit regularization through weight decay is required in this experiment. However, this parame-

ter can be adjusted if necessary to prevent overfitting in future studies.

- Dropout Rate: A dropout rate of 0.5 is applied to the embedding layer. This is a standard regularization technique that helps prevent overfitting by randomly "dropping" connections during training, encouraging the model to learn more robust features.

2. Graph Convolutional Network (GCN) Configuration:
The GCN component of the model is responsible for capturing relational dependencies between words in a sentence. The hyperparameters specific to the GCN are as follows:

    - GCN Hidden Dimension: The hidden dimension of the GCN is set to 300. This value is selected to effectively process graph-structured data while maintaining a manageable computational burden. This dimensionality allows the model to capture sufficient relational information without overloading the model with excessive complexity.

3. Encoder and Optimizer Configuration
At the core of our model, we utilize the BERT-base-uncased encoder (version 5), which is known for its ability to generate context-aware word embeddings. To fine-tune BERT for the specific task at hand, we use the AdamW optimizer [57], a variant of the Adam optimizer that introduces weight decay correction to mitigate the issues of traditional adaptive learning rate methods. The learning rate for BERT fine-tuning is set to 2×10^{-5} to avoid disrupting the pre-trained parameters while ensuring effective adaptation to the new task.
4. Training Configuration
The model is trained over a total of 100 epochs, with a batch size of 16. This training setup is designed to ensure sufficient exposure to the training data while minimizing overfitting risks. The hyperparameters γ and δ, which control the influence of relation constraints, are set to 0.1 and 0.01, respectively, to fine-tune the model's sensitivity to relational information.
5. Evaluation Metrics
To rigorously evaluate the performance of the Fpa-GCN model, we use three standard evaluation metrics: Precision (P), Recall (R), and macro-F1 (F1) score. These metrics are chosen because they provide a comprehensive assessment of the model's effectiveness in identifying and classifying aspect-sentiment pairs across multiple benchmark datasets.

The implementation details of the Fpa-GCN model are carefully crafted to align with the best practices in the field, ensuring that the experimental results are both reliable and indicative of the model's capabilities in addressing the ASTE task.

4.4 Main results

This subsection presents the primary outcomes derived from our experiments on ASTE tasks for data versions $\mathcal{D}_1$ and $\mathcal{D}_2$, as showcased in Tables 3 and 4, respectively. The analysis of the results contained within these tables yields pivotal insights into the performance of various models, particularly emphasizing the superiority of Pre-trained Language Models (PLMs) over traditional word2vector approaches in executing ASTE tasks.

A noteworthy highlight from our findings is the exceptional performance of the Fpa-GCN model, which consistently outperforms existing state-of-the-art models across all datasets encompassed within $\mathcal{D}_1$ and $\mathcal{D}_2$. For instance, within $\mathcal{D}_1$, Fpa-GCN demonstrates significant F1 score improvements, marking enhancements of 1.31%, 0.99%, 0.11%, 3.7%, and 2.15% over the SOTA on four distinct datasets. Analogously, Fpa-GCN achieves remarkable F1 score uplifts of 1.66%, 0.08%, 4%, and 3.33% compared to the SOTA within the context of $\mathcal{D}_2$. These results unequivocally illus-

Table 3 The performance of Fpa-GCN on $\mathcal{D}_1$

Models	Restaurant14			Laptop14			Restaurant15			Restaurant16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Peng-two-stage+IOG©[56]	58.89	60.41	59.64	48.62	45.52	47.02	51.70	46.06	48.71	59.25	58.09	58.67
GTS-CNN©[56]	70.79	61.71	65.94	55.93	47.52	51.38	60.09	53.57	56.64	62.63	66.98	64.73
GTS-BiLSTM©[56]	67.28	61.91	64.49	59.42	45.13	51.30	63.26	50.71	56.29	66.07	65.05	65.56
GTS-BERT©[56]	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	68.58	66.60	67.58
EMC-GCN † [23]	70.92	71.49	71.20	58.96	54.31	56.54	54.99	61.46	58.04	65.74	72.66	69.03
$S3̂E2̂$©	69.08	64.55	66.74	59.43	46.23	52.01	61.06	56.44	58.66	71.08	63.13	66.87
Dual-MRC©[15]	–	–	70.32	–	–	55.58	–	–	57.21	–	–	67.40
BMRC	–	–	70.01	–	–	57.83	–	–	58.74	–	–	67.49
Fpa-GCN	72.66	70.41	**72.51**	58.38	56.72	**57.94**	60.48	63.06	**62.44**	69.45	71.65	**71.18**

Table 4 The performance of Fpa-GCN on $\mathcal{D}_2$

Models	Restaurant14			Laptop14			Restaurant15			Restaurant16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
RINANTE+♯	31.42	39.38	34.95	21.71	18.66	20.07	29.88	30.06	29.97	25.68	22.30	23.87
Peng-two-stage♯	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	46.96	64.24	54.21
CMLA+♯	39.18	47.13	42.79	30.09	36.92	33.16	34.56	39.84	37.01	41.34	42.10	41.72
MuG-BERT©	68.40	67.64	68.00	58.30	52.21	55.06	60.65	54.12	57.10	66.26	67.39	66.74
BART-ABSA†	65.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
EMC-GCN †	70.35	73.14	71.72	61.48	55.45	58.31	56.33	63.30	59.61	62.46	72.32	67.03
BMRC†	75.61	61.77	67.99	70.55	48.98	57.82	68.51	53.40	60.02	71.20	61.08	65.75
GTS-BERT†	68.09	69.54	68.81	59.40	51.94	55.42	59.28	57.93	58.60	68.32	66.86	67.58
Fpa-GCN	72.49	74.30	**73.38**	59.73	55.82	**58.77**	63.25	64.18	**64.02**	66.38	73.56	**70.95**

trate Fpa-GCN's dominance, underscoring its unparalleled efficacy across the entirety of the datasets under consideration.

The performance metrics delineated in Table 3 reveal that, within $\mathcal{D}_1$, Fpa-GCN attains an F1 score of 72.51% on the Restaurant14 dataset, a notable leap from the closest competitor, EMC-GCN, which scored 71.20%. Similar trends are observable across other datasets within $\mathcal{D}_1$, wherein Fpa-GCN consistently leads in performance metrics, thereby establishing its pre-eminence. Furthermore, the examination of Table 4 delineates Fpa-GCN's superior performance within $\mathcal{D}_2$. Notably, Fpa-GCN secures an F1 score of 73.38% on the Restaurant14 dataset of $\mathcal{D}_2$, surpassing the next best model, EMC-GCN, which achieved an F1 score of 71.72%. This trend of Fpa-GCN outpacing its contenders persists across the datasets in $\mathcal{D}_2$, further cementing its status as the leading model for ASTE tasks.

Our comprehensive analysis underscores two principal observations: firstly, the evident supremacy of PLM-based approaches over conventional word2vector models for ASTE tasks, and secondly, the consistent outperformance of Fpa-GCN across all evaluated datasets, validating its SOTA status.

5 Model analysis

5.1 Ablation study

The Fpa-GCN model incorporates multiple innovative modules designed to enhance its performance in the ASTE task. Table 5 presents a detailed analysis of ablation studies, evaluating the contribution of individual components to the model's overall effectiveness. These studies were conducted on four datasets: Restaurant14, Laptop14, Restaurant15, and Restaurant16. The baseline Fpa-GCN model achieves the highest performance across all datasets. However, removing critical components-such as the Information Fusion module, Gating Mechanism, Biaffine Attention, or Linguistic Features-results in varying degrees of performance degradation. Notably, the removal of Biaffine Attention causes the most significant performance drop, especially on the Restaurant16 dataset (↓5.52), underscoring its crucial role in the model's success. This systematic ablation analysis highlights the importance of these components, and a detailed examination of their individual contributions is provided in the following sections.

5.1.1 Effect of information fusion

Information Fusion, a cornerstone of our proposed methodology, is meticulously designed to amalgamate and synthesize global information extracted from two primary sources: BERT embeddings and features derived from a Fpa-GCN. This integration is pivotal in refining the semantic understanding of texts by amalgamating the nuanced contextual insights from BERT with the structured relational informationcaptured through Fpa-GCN, thereby enriching the model's comprehension of semantic nuances while mitigating the adverse effects of parsing inaccuracies.

Table 5 F1 scores of ablation study on D2

	Restaurant14	Laptop14	Restaurant15	Restaurant16
Fpa-GCN	73.38	58.77	64.02	70.95
w/o Information Fusion	71.89 (↓1.49)	58.51 (↓0.26)	59.93 (↓4.09)	67.15 (↓3.80)
w/o Gating Mechanism	72.06 (↓1.32)	58.66 (↓0.11)	62.23 (↓1.79)	68.56 (↓2.39)
w/o Biaffine attention	72.12 (↓1.26)	57.12 (↓1.65)	61.52 (↓2.50)	65.43 (↓5.52)
w/o Linguistic Features	71.12 (↓2.26)	58.18 (↓0.59)	60.62 (↓3.40)	66.37 (↓4.58)

Note: The "w/o" denotes the abbreviation for without

Table 6 The contribution of information fusion to Fpa-GCN for ASTE task on $\mathcal{D}1$ and $\mathcal{D}2$

Versions	Models	Restaurant14			Laptop14			Restaurant15			Restaurant16		
		P	R	F1	P	R	F1	P	R	F1	P	R	F1
D1	Fpa-GCN	72.66	70.41	72.51	58.38	56.72	57.94	60.48	63.06	62.44	69.45	71.65	71.18
	w/o Information Fusion	71.11	71.80	71.43	59.22	54.55	56.83	55.26	61.67	58.19	65.92	72.88	69.18
D2	Fpa-GCN	72.49	74.30	73.38	59.73	55.82	58.77	63.25	64.18	64.02	66.38	73.56	70.95
	w/o Information Fusion	70.47	73.33	71.89	61.77	55.70	58.51	56.62	63.51	59.93	62.73	72.59	67.15

Note: The "w/o" denotes the abbreviation for without

To empirically validate the efficacy of information fusion, we conducted a comprehensive series of experiments on two distinct dataset versions, namely $\mathcal{D}1$ and $\mathcal{D}2$, across four datasets. The results of these experiments are systematically presented in Table 6. The comparative analysis focuses on the performance metrics of Precision (P), Recall (R), and the F1-score (F1) for the tasks of ASTE across the datasets Restaurant14, Laptop14, Restaurant15, and Restaurant16.

The empirical data gained from these experiments underscores a significant performance decrement in scenarios devoid of Information Fusion modular. Conversely, the incorporation of this technique manifests a pronounced enhancement in functionality, as evidenced by improved metrics. Figure 5 provides a compelling visual representation of the softmax probability distributions for each label, contrasting the outputs of the Fpa-GCN model with and without the application of Information Fusion. The horizontal axis of both heat maps denotes the position of the data within the input sequence, showcasing the model's prediction for each token, while the vertical axis represents the ten different classes under consideration.

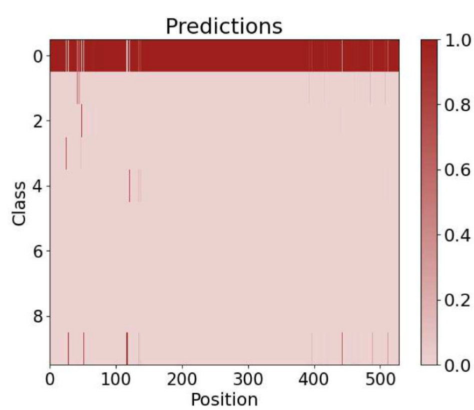
In Fig. 5(a), the Fpa-GCN model with Information Fusion shows a denser distribution of darker shades, particularly in the upper sections of the map, which correspond to the higher probability predictions. This density indicates a strong confidence in the model's predictions across a majority of the classes. In contrast, the heat map shown in Fig. 5(b), which represents the model's performance without Information Fusion, exhibits sparser dark regions, reflecting a less decisive model output. The visualization presented above clearly indicates a significant difference in the level of confidence between two models. The model that incorporates information fusion exhibits a more uniform and stronger prediction capability overall.

A detailed examination of the heat maps, particularly with data from the laptop14 subset of the $\mathcal{D}1$ dataset, reveals two critical insights. Firstly, a substantial portion of the text was identified as extraneous filler content, which could potentially obfuscate the model's predictive accuracy. Secondly, and more importantly, the integration of Information Fusion markedly improved the model's confidence in correctly predicting the relevant labels. This not only signifies a heightened accuracy in label prediction but also indicates a reduced susceptibility to noise and misinterpretations stemming from irrelevant text segments. The data presented in Table 6 and the visual insights from Fig. 5 collectively highlight the indispensable role of Information Fusion in augmenting the precision and robustness of semantic text analysis models.

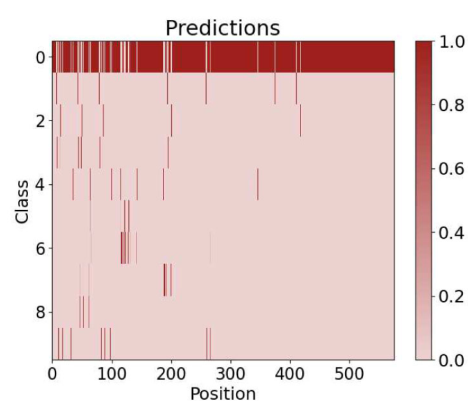
5.1.2 Effect of linguistic feature

As shown in Table 5, the results of our ablation experiments, where all linguistic features were removed, demonstrate a

Fig. 5 Visualizing softmax probability distributions across various classes



(a)Fpa-GCN



(b)w/o Information Fusion

significant decrease in the F1 scores across the four benchmark datasets. Specifically, the scores dropped by 2.26, 0.59, 3.40, and 4.58, respectively. This stark decline underscores the critical role of linguistic features in enhancing the Fpa-GCN model's performance. Figure 6 further outlines the F1 score percentages for different configurations of the Fpa-GCN model across four benchmark datasets, particularly under the $\mathcal{D}1$ version. Each bar represents a unique combination of the model integrating semantic features from BERT with a specific type of linguistic feature. The bar patterns correspond to the different linguistic features integrated with the base model: F^G_{ba+sdt} employs syntactic dependency trees (R_{sdt}), F^G_{ba+psc} incorporates part-of-speech tags coupled with syntactic chunks (R_{psc}), F^G_{ba+tbd} is based on the integration of top-down body language features (R_{tbd}), and F^G_{ba+rpd} utilizes refined positional dependencies (R_{rpd}).

The results displayed indicate that the integration of R_{psc} and R_{sdt} with the Fpa-GCN model yields the most pronounced enhancement in performance across the four datasets, particularly in the Restaurant14 and Restaurant15 benchmarks. These results highlight the pivotal role that specific linguistic features play in refining the semantic representation derived from BERT, thereby leading to superior model efficacy.

Conversely, while the inclusion of R_{tbd} and R_{rpd} linguistic features does not ascend to the same level of effectiveness as R_{psc} and R_{sdt}, the improvement in the F1 scores over the base model is still noteworthy, emphasizing their contributory value to the model's overall capability in ASET. These experimental findings verify the hypothesis that the careful integration of linguistic features can significantly enhance the performance of semantic feature-based models.

5.1.3 Effect of gating mechanism

Assessing the limited occurrence of crucial words in sentences, the gating mechanism in natural language processing proves pivotal. It efficiently reduces background noise, allowing the model to focus on key words, thereby significantly improving overall performance and accuracy. Table 5 encapsulates the results of a comparative analysis designed to empirically validate the efficacy of the gating mechanism within the Fpa-GCN framework for the ASTE task. The evaluation was conducted on the $\mathcal{D}2$ dataset, specifically on the subsets of Restaurant14 and Laptop14, among others.

An examination of the performance metrics presented in the table reveals a detectable improvement in the Fpa-GCN model's outcomes when the gating mechanism is operational. The model equipped with the gating mechanism consistently demonstrates superior F1 scores compared to its counterpart without it. For instance, in the Restaurant14 dataset, the Fpa-GCN model with gating achieves an F1 score of 73.38%, as opposed to 72.06% when the gating mechanism is excluded, underscoring its key role in the model's ability. Similarly, in the Laptop14 dataset, the Fpa-GCN model with gating yields an F1 score of 58.77%, illustrating a slight yet impactful enhancement over the 58.66% scored by the model without the gating feature. Although the gating mechanism's influence is relatively more modest in the Laptop14 dataset, it still contributes to the overall effectiveness of the model. The increased F1 scores across various datasets confirm that the gating mechanism is capable of filtering out irrelevant information. This suggests that the model is able to focus on important words and improve its ability to perform the ASTE task with greater accuracy and dependability.

5.1.4 Effectiveness comparison between information fusion and gating mechanism modulars

The results of all ablation experiments conducted on the dataset $\mathcal{D}2$ are systematically presented in Fig. 7, which details the performance impact of Information Fusion and the Gating Mechanism modules when incorporated within the Fpa-GCN framework. This comparative analysis is visually represented through bar graphs, depicting the model's Precision, Recall, and F1 score across four different datasets: Restaurant14, Laptop14, Restaurant15, and Restaurant16. Each group of bars within the graph correlates to a specific dataset, showcasing three distinct configurations: the Fpa-GCN model with all features integrated (denoted by the red horizontal lines), the model sans Information Fusion (blue diagonal lines), and the model without the gating mechanism (green dots pattern). The contrast between these configurations provides a clear comparative view of their effectiveness.

From the graph, it is discernible that the full Fpa-GCN model tends to show the highest F1 scores, indicating that the synergetic effect of both Information Fusion and Gating Mechanism modules contributes positively to the model's overall performance. The reduction in F1 scores when either module is omitted is a testament to the significance of each module in enhancing the model's capabilities. The comparison across datasets further demonstrates the consistent benefits provided by these modules, regardless of the dataset's domain. Notably, the absence of the Gating Mechanism appears to impact the Precision and Recall metrics slightly differently depending on the dataset, suggesting that the module's role may be influenced by the specific linguistic characteristics or the domain of the dataset. Nonetheless, the overall trend indicates that both Information Fusion and the Gating Mechanism are essential components that improve the accuracy and reliability of the Fpa-GCN model in the Aspect Sentiment Triplet Extraction task on $\mathcal{D}2$ dataset.

Each group of bars within the graph correlates to a specific dataset, showcasing three distinct configurations: the Fpa-GCN model with all features integrated (denoted by the red

Fig. 6 F1 on four benchmarks with each branch GCN on $\mathcal{D}1$

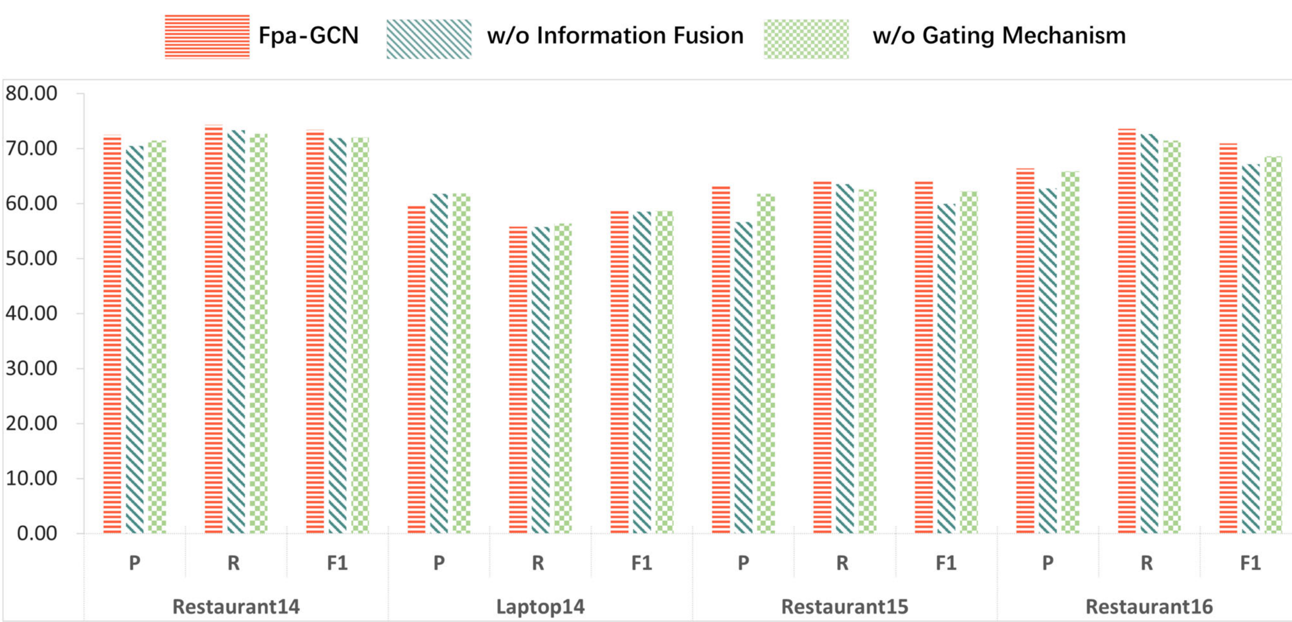


Fig. 7 Detailed findings from the ablation study of Fpa-GCN on dataset $\mathcal{D}2$

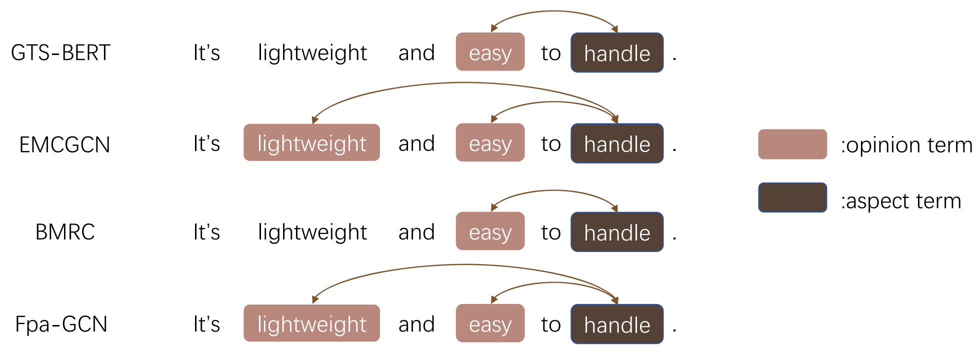


Fig. 8 Different models outputs for a given sentence

checkered pattern), the model sans information fusion (blue dots pattern), and the model without the gating mechanism (gray stripes pattern). The contrast between these configurations provides a clear comparative view of their effectiveness.

5.2 Case study

5.2.1 Impact of spatial distance on aspect-opinion pairing: a model comparison

Figure 8 illustrates the outputs of four models—GTS-BERT, EMCGCN, BMRC, and Fpa-GCN—on a given sentence, where each model attempts to match the aspect term "handle" with the corresponding opinion terms "lightweight" and "easy." The golden opinion term "lightweight" poses a challenge for identification due to its relative distance from "handle" in the sentence. GTS-BERT and BMRC models fail to correctly predict the triplet ('handle', 'lightweight', 'positive'), likely due to their weaker ability to capture sentiment relations or to insufficiently leverage linguistic features. Both GTS-BERT and Fpa-GCN utilize BERT as a pre-trained model, while EMCGCN and Fpa-GCN incorporate Graph Convolution Networks (GCN), allowing these models to effectively capture relational structures. BMRC, on the other hand, introduces a novel approach that combines different strategies. By comparing these three models with Fpa-GCN, we can more clearly highlight the advantages of our model. Specifically, Fpa-GCN demonstrates a superior capacity to establish meaningful connections through sentiment relations and linguistic cues, leading to more accurate triplet identification.

5.2.2 Error analysis and model improvement

In this chapter, we conduct a comprehensive analysis of the model's errors to better understand the model's strengths and weaknesses and gain a deeper understanding of its areas for improvement. As illustrated in Fig. 9, we randomly selected an incorrectly extracted triplet from the dataset for further examination. Specifically, the input provided to the model was: "A restaurant that doesn't try to do anything except serve great food with great service in a pleasant atmosphere." The corresponding incorrect triplet extracted was: restaurant, doesn't try, negative. This error suggests that the model misinterpreted "doesn't try" as a negative sentiment expression. However, based on the sentence semantics, this phrase actually describes the restaurant's behavior rather than expressing an emotional stance.

The following potential causes of this issue are identified:

1. Misinterpretation of Sentence Structure: In the sentence "doesn't try to do anything except serve great food with

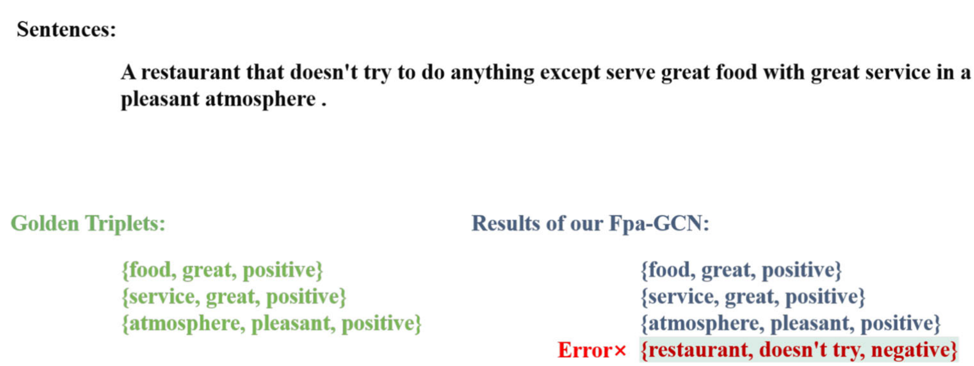


Fig. 9 Information taken by our Fpa-GCN. Errors in comparison with gold labels are highlighted and annotated

great service in a pleasant atmosphere," the model may have erroneously classified "doesn't try" as a negative sentiment expression. In fact, the intention behind this phrase is to highlight the restaurant's focus and effort, not to provide a negative evaluation.

2. Insufficient Contextual Analysis of Sentiment Terms: While "doesn't try" might imply negative sentiment in some contexts, in this case, the phrase does not directly relate to any specific aspect of the restaurant (such as food, service, or atmosphere). The model failed to distinguish between descriptive content and actual sentiment pairs.
3. Bias in Training Data Annotations: If similar instances exist in the training data where "doesn't try" is labeled as a negative sentiment, the model may tend to apply this erroneous inference to new sentences. Therefore, it is important to verify whether such misannotations exist in the training data.
4. Limitations of the Sentiment Recognition Strategy: The model's sentiment recognition strategy may not properly handle negation words, such as "doesn't," leading to misclassification. Particularly in sentences with complex semantic structures, the model may incorrectly associate "doesn't try" with negative sentiment.

To address these issues, we propose the following improvements:

1. Enhance Contextual Understanding: Improve the model's ability to interpret context in longer sentences, ensuring it can accurately distinguish between emotional expressions and merely descriptive content, especially in complex sentiment judgments.
2. Refine Data Annotation Guidelines: To prevent misannotations, we recommend more precise and detailed labeling of sentiment relationships in the training data, ensuring that sentiment expressions are clearly delineated and not oversimplified.
3. Improve Handling of Negation: The model's sentiment inference strategy should be optimized to better handle negation, particularly in contexts where multiple layers of meaning co-exist with sentiment expressions. This will ensure that negations are not mistakenly classified as indicators of negative sentiment.

By implementing these improvements, we believe the model will achieve greater accuracy in sentiment triplet extraction, reduce errors, and ultimately enhance the performance of Fpa-GCN in real-world applications.

5.2.3 Real-world application of Fpa-GCN

This section presents an in-depth case study to demonstrate the ability of our model in extracting triplets for the ASTE task and further illustrates how the model operates in real-world applications. To better present this process, we utilized five comments about electronic device online shopping, generated by a large language model, as raw data. Specifically, as shown in Fig. 10(a), the initial raw data input was: "The keyboard is too slick. ... I love the glass touchpad." After processing with the Fpa-GCN model, we successfully extracted the following triplets: ("keyboard", "slick", "NEG") ... ("touchpad", "love", "POS").

Figure 10(b) shows a visualized triplet extraction network, with the triplets displayed representing the results obtained after model processing.

In this visualization, the aspect terms are represented by light red nodes, highlighting their role as the key features in the sentence. For example, the aspect terms "keyboard," "website," "MacBook Pro," "warranty," and "touchpad" are displayed as light red nodes, each linked to corresponding opinion nodes. The opinion nodes are depicted as light green, emphasizing their function as the evaluations or attributes associated with the aspects. For instance, "slick," "freezes," "fast," "bad," and "love" are the opinion terms linked to the respective aspects in the triples.Sentiment polarity is represented by grey, orange, and purple nodes, each corresponding to different sentiment categories. Grey nodes represent positive sentiment (e.g., "POS"), orange nodes represent negative sentiment (e.g., "NEG"), and purple nodes represent neutral sentiment (e.g., "NEU").

The network graph provides a clear view of how each aspect, opinion, and sentiment are linked. For example, the aspect "keyboard" is associated with the opinion "slick" and the negative sentiment "NEG." Similarly, the aspect "MacBook Pro" is connected to the opinion "fast" and the positive sentiment "POS."

The analysis of the extracted triplets reveals that customers express positive sentiments toward the "website," "MacBook Pro," and "touchpad," suggesting that the retailer should maintain the current product design and service quality. In contrast, the sentiment for "warranty" is neutral, indicating potential areas for improvement. For the "keyboard," however, significant negative sentiment is observed, signaling a critical issue that could impact both sales and customer satisfaction. This issue warrants immediate attention and product improvement.

This analysis not only highlights the efficiency of the Fpa-GCN model in extracting sentiment triplets but also underscores its practical advantages. Using the Fpa-GCN model, retailers can accurately identify the sentiment direction in customer feedback, allowing them to pinpoint specific

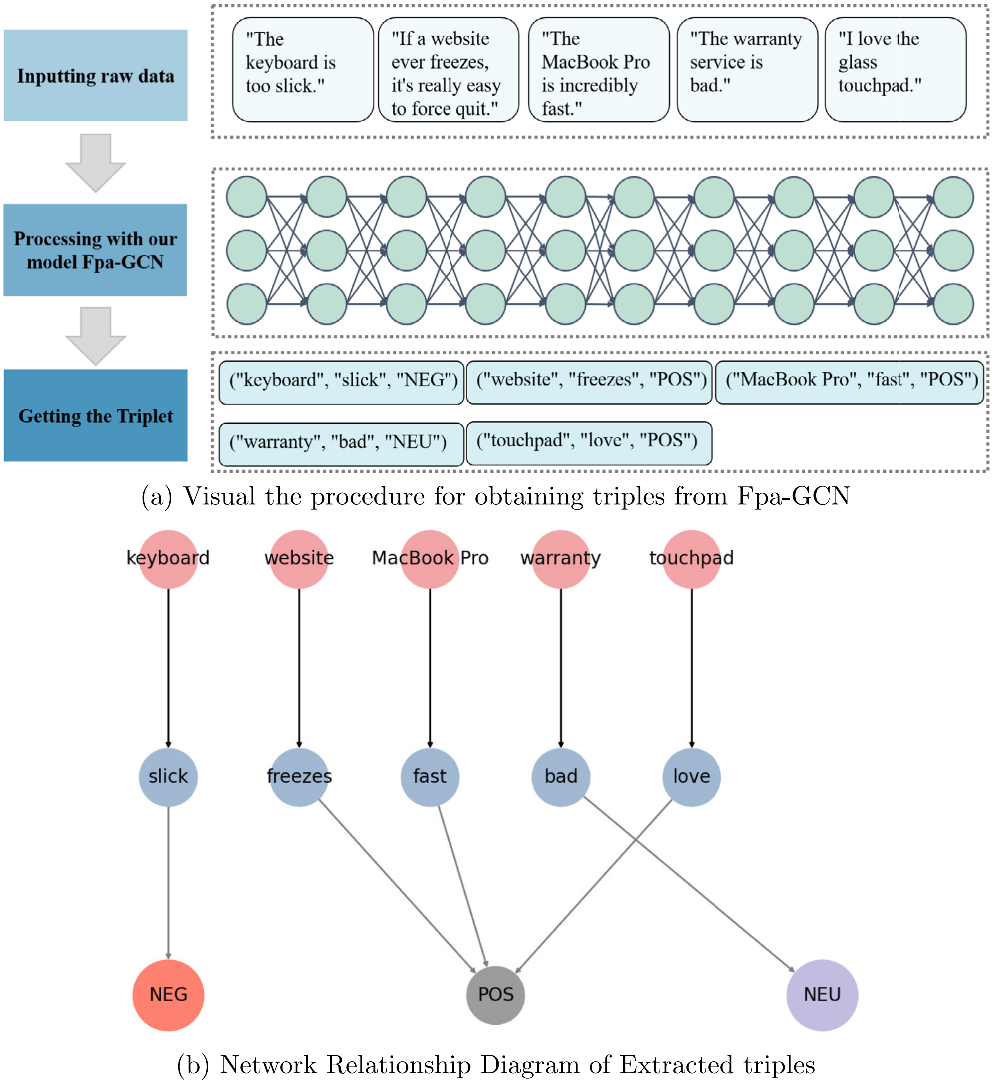


Fig. 10 Visualization of a real-world application case

product features in need of improvement. Compared to traditional methods, Fpa-GCN excels in industrial and real-world applications, handling more complex and diverse emotional expressions. Particularly in large-scale data processing scenarios, Fpa-GCN provides accurate, real-time feedback, enabling retailers to make targeted, actionable improvements.

5.3 Further analysis

5.3.1 Model simplification and computational efficiency

This section presents an evaluation of model simplification and distributed training techniques to assess their effectiveness in improving computational efficiency while maintaining model performance.

To validate the effectiveness of model simplification methods, we systematically compared the performance of lightweight models based on parameter count and evaluation metrics (e.g., F1 score). The experimental results, presented in Table 7, show that using TinyBERT as the pre-trained model significantly reduced its parameter count from 112M (original BERT) to 16.85M, with only a marginal 1%-2% decrease in F1 score. This result demonstrates that the model maintains stable performance while significantly lowering hardware and computational resource requirements. Such a favorable trade-off between parameter count and performance underscores the practicality and research significance of Fpa-GCN in resource-constrained scenarios.

Additionally, the experimental results confirm the model's adaptability and efficiency in specific tasks which enhances deployment flexibility and provides critical insights into optimizing the design and application of models.

Table 7 Comparison of pretrained models: the term ‘parameter‘ refers to the total number of parameters in the entire Fpa-GCN when utilizing the pretrained model

Pretrained Model	Parameters	VRAM	F1 Score On Different Datasets			
			Restaurant14(D2)	Laptop14(D2)	Restaurant15(D2)	Restaurant16(D2)
Bert-Base-Uncased	112M	7GB	73.38	58.77	64.02	70.95
TinyBERT	16.85M	5GB	71.85 (↓ 1.53)	57.16 (↓ 1.61)	62.17 (↓ 1.85)	69.53 (↓ 1.42)

‘VRAM‘ refers to the amount of Video RAM (GPU memory) consumed during training

Table 8 summarizes the changes in training speed, VRAM usage, and F1 scores across various datasets under different numbers of GPUs. The results demonstrate that distributed training significantly accelerates training speed while maintaining model prediction performance, as measured by F1 scores. With an increase in the number of GPUs from 1 to 3, the training speed improves markedly. For instance, in a single-GPU setup, the average epoch time is 24 seconds, which decreases to 15 seconds and 10 seconds for dual-GPU and triple-GPU setups, respectively—yielding speedups of approximately 37.5% and 58.3%. These findings highlight the effectiveness of distributed training in significantly reducing training time by partitioning training tasks across multiple GPUs on our Fpa-GCN. Additionally, the experiments reveal that distributed training has a negligible impact on model performance. As shown in Table 8, the F1 score variation across different datasets is less than 1%, indicating that allocating computational tasks via distributed strategies does not substantially affect model prediction accuracy. This result reinforces the practical value of distributed training, which enhances training efficiency without compromising performance.

For example, in the Restaurant14 dataset, the F1 score increased from 73.38 in the single-GPU setup to 73.80 and 73.83 in the dual-GPU and triple-GPU setups, representing improvements of 0.42 and 0.45, respectively. Similarly, for the Restaurant16 dataset, the F1 score in the triple-GPU setup (71.50) slightly exceeded that of the single-GPU setup (70.95), further validating the robustness of distributed training. The rapid training speed not only shortens research cycles for researchers but also facilitates real-time or near-real-time model updates in industrial applications.

These findings emphasize the potential of Fpa-GCN in optimizing machine learning models for practical deployment, especially in scenarios where resources are constrained.

5.3.2 Evaluating model robustness and scalability

To ensure that our model not only performs well on the ASTE task but also demonstrates good task scalability and robustness, we conducted comprehensive testing. As shown in Table 9, the dataset, proposed by [58], is designed specifically for the “quadruple extraction task in dialogue scenarios.” It is built using data from Weibo, the largest social media platform in China, known for its diverse and dynamic conversational content. This novel dataset comprises 9 million posts and comments from 100 verified digital bloggers, forming a tree-structured, multi-threaded, and multi-turn dialogue framework. Following extensive preprocessing and rigorous cleaning, 1,000 high-quality dialogues were selected to ensure the dataset accurately reflects real-world conversational dynamics.

These dialogues were annotated by crowdsourced workers trained following the SemEval ABSA guidelines. To ensure annotation quality, linguistics and computer science experts reviewed the annotations. Cross-checking and validation were performed using automated rules, achieving high inter-annotator agreement with a Cohen’s Kappa score of 0.86.

We also visualized this dataset using a word cloud, as shown in Fig. 11, where words with higher frequencies appear in larger font sizes. A closer look reveals that words related to targets, aspects, and opinions appear infrequently in the dialogue data, while more common words such as “the,” “I,” “and,” and “it” dominate. This observation reflects real-life situations, where such words do not always hold prominent positions in conversations. We believe that our

Table 8 Impact of GPU count on model performance and training speed: ‘GPUs‘ refers to the number of GPUs utilized during training, while ‘VRAM‘ refers to the amount of Video RAM (GPU memory) consumed during training

GPUs	Speed	VRAM	F1 Score On Different Datasets			
			Restaurant14(D2)	Laptop14(D2)	Restaurant15(D2)	Restaurant16(D2)
1	24s/Epoch	7GB	73.38	58.77	64.02	70.95
2	15s/Epoch	16GB	73.80 (↑0.42)	59.20 (↑0.43)	64.50 (↑0.48)	70.70 (↓0.25)
3	10s/Epoch	23GB	73.83 (↑0.45)	58.66 (↓0.11)	64.70 (↑0.68)	71.50 (↑0.55)

Table 9 statistics for the DiaASQ dataset English version

		Dialogue			Items			Pairs			Quadruples		
		Dia.	Utt.	Spk.	Tgt.	Asp.	Opi.	$Pair_{t-a}$	$Pair_{t-o}$	$Pair_{a-o}$	Quad.	Intra.	Cross.
DIaASQ-EN	Total	1,000	7,452	4,991	8,264	6,434	6,933	5,894	7,432	4,994	5,514	4,287	1,227
	Train	800	5,947	3,986	6,613	5,109	5,523	4,699	5,931	3,989	4,414	3,442	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

model's performance on this complex and noisy social media dataset will provide stronger evidence for its generalizability.

As shown in Table 10, we compared the model's performance on the Span Match and Pair Extraction tasks with previous state-of-the-art models, using the DiaASQ English version of the dataset (DiaASQ-EN).

The CRF-ExtractClassify model [59] introduces Conditional Random Fields (CRF) combined with classification techniques to address issues of implicit aspects and viewpoints in *Aspect-Category-Opinion-Sentiment* (ACOS) analysis, pioneering new research directions and datasets. The SpERT model [60], based on attention mechanisms, simultaneously extracts entities and relationships. Its lightweight reasoning framework integrates BERT embeddings and employs intra-sentence negative sampling to enhance span detection and relation classification efficiency. The Span-ASTE method [61] utilizes a dual-channel span pruning technique that integrates aspect term extraction (ATE) and opinion term extraction (OTE), significantly improving computational efficiency and the ability to predict semantic consistency. Another study [62] redefines the *Aspect Sentiment Quadruple Prediction* (ASQP) task as a unified paraphrase generation problem and proposes a paraphrase-based generative framework that predicts sentiment quadruples using semantic information, addressing the error propagation issues of traditional pipeline methods. DiaASQ [58] focuses on the task of dialogue sentiment quadruple extraction by constructing a bilingual dataset in Chinese and English and proposing a baseline model based on word-pair relationship modeling, bridging gaps in viewpoint mining and conversational sentiment analysis. IFusionQuad [63] introduces an innovative framework specifically designed for quadruplet extraction in dialogue scenarios (DiaASQ task). Its key contributions are twofold. First, the framework incorporates the CloBlock module, which employs a context-aware weighting mechanism to enhance the accuracy of feature representation, enabling the effective capture of complex semantic relationships within dialogues. Second, it features the InteractiveNet Fusion module, which integrates multidimensional dialogue characteristics—including threads, speakers, and replies—to comprehensively improve the understanding of dialogue semantics.These studies provide innovative methodologies and solid theoretical support for different sentiment analysis tasks, laying the foundation for advancements in the field.

As presented in Table 10, the Fpa-GCN model exhibited exceptional performance across two tasks on the dialogue dataset (DiaASQ-EN), despite not being originally designed for dialogue data. In the Span Match task, the F1 score for Aspect reached 74.92, the highest among all models. For Opinion, the F1 score was 62.87, only 0.61% lower than the top score of 63.48 achieved by the IFusionQuad model. In the Pair Extraction task, while the IFusionQuad model achieved the highest F1 score of 51.94, the Fpa-GCN model secured the second-best score of 48.32, substantially outperforming most other comparative models.

Fig. 11 Visual word cloud for a dialogue

Table 10 Results of the Span Matc and Pair Extraction tasks for English datasets

		Span Match (F1)		Pair Extraction (F1)
		A	O	A-O
DIaASQ-EN	CRF-Extract-Classify	71.71	47.90	19.21
	SpERT	74.65	54.17	23.64
	ParaPhrase	/	/	30.78
	Span-ASTE	/	/	45.90
	DiaASQ	74.71	60.22	44.27
	IFusionQuad	74.23	**63.48**	**51.94**
	Our Fpa-GCN	**74.92**	62.87	48.32

'A, O' stands for Aspect and Opinion, respectively

Notably, although the Fpa-GCN model was not specifically designed for dialogue scenarios, it surpassed all models in the Span Match: Aspect task within this dataset. This result underscores the model's remarkable generalization capabilities across diverse tasks and data distributions. Such performance highlights the potential of the Fpa-GCN model in handling complex scenarios and lays a solid foundation for broader applications.

These results demonstrate our model's robustness and scalability across various tasks. Additionally, they highlight the model's ability to handle diverse and informal text data, such as social media posts and user-generated comments, further emphasizing its universality and robustness.

5.3.3 Hyperparameter sensitivity analysis

This section presents a systematic sensitivity analysis of key hyperparameters to explore their impact on model performance. Hyperparameter optimization plays a critical role in enhancing the effectiveness and generalizability of machine learning models. In this analysis, we focus on the effect of training epochs and the hidden dimensions of the Graph Convolutional Network (GCN_dim), which are fundamental to the performance of the Fpa-GCN model. The results, summarized in Table 11, illustrate the F1 scores achieved by the model across four datasets (Restaurant14, Laptop14, Restaurant15, and Restaurant16) under different configurations of training epochs and GCN_dim. For enhanced clarity and interpretability, these results are also visualized in Fig. 12.

The analysis indicated that the model achieved optimal performance at 100 epochs, beyond which performance declined, suggesting the onset of overfitting. In terms of GCN_dim, the highest F1 scores across all datasets were obtained with a dimension of 300. Smaller dimensions led to insufficient feature representation, while larger dimensions introduced redundancy, both of which adversely affected performance.

These findings deepen our understanding of the interplay between hyperparameters and model performance, offering valuable guidance for future hyperparameter selection. Additionally, they contribute to enhancing the reproducibility and applicability of the model.

6 Conclusions and future work

This study presents the Feature-riched Prediction-Aware Graph Convolutional Network (Fpa-GCN), an end-to-end, advanced framework tailored to address the complexities of aspect sentiment triplet extraction. By leveraging multi-branch GCNs, the Fpa-GCN model integrates diverse linguistic features with high efficacy and employs a gating mechanism to reduce noise interference. This innovative approach harmonizes global and local information, thereby enhancing robustness and anti-interference capabilities. Furthermore, it diverges from previous methodologies that often overlooked the intricate interplay among triplet elements and relied on less coherent strategies for constructing contextual representations.

Our extensive evaluation on publicly available benchmark datasets demonstrates the superior performance of the Fpa-GCN model, supported by comprehensive comparative analyses with baseline models. Key contributions of this study include:

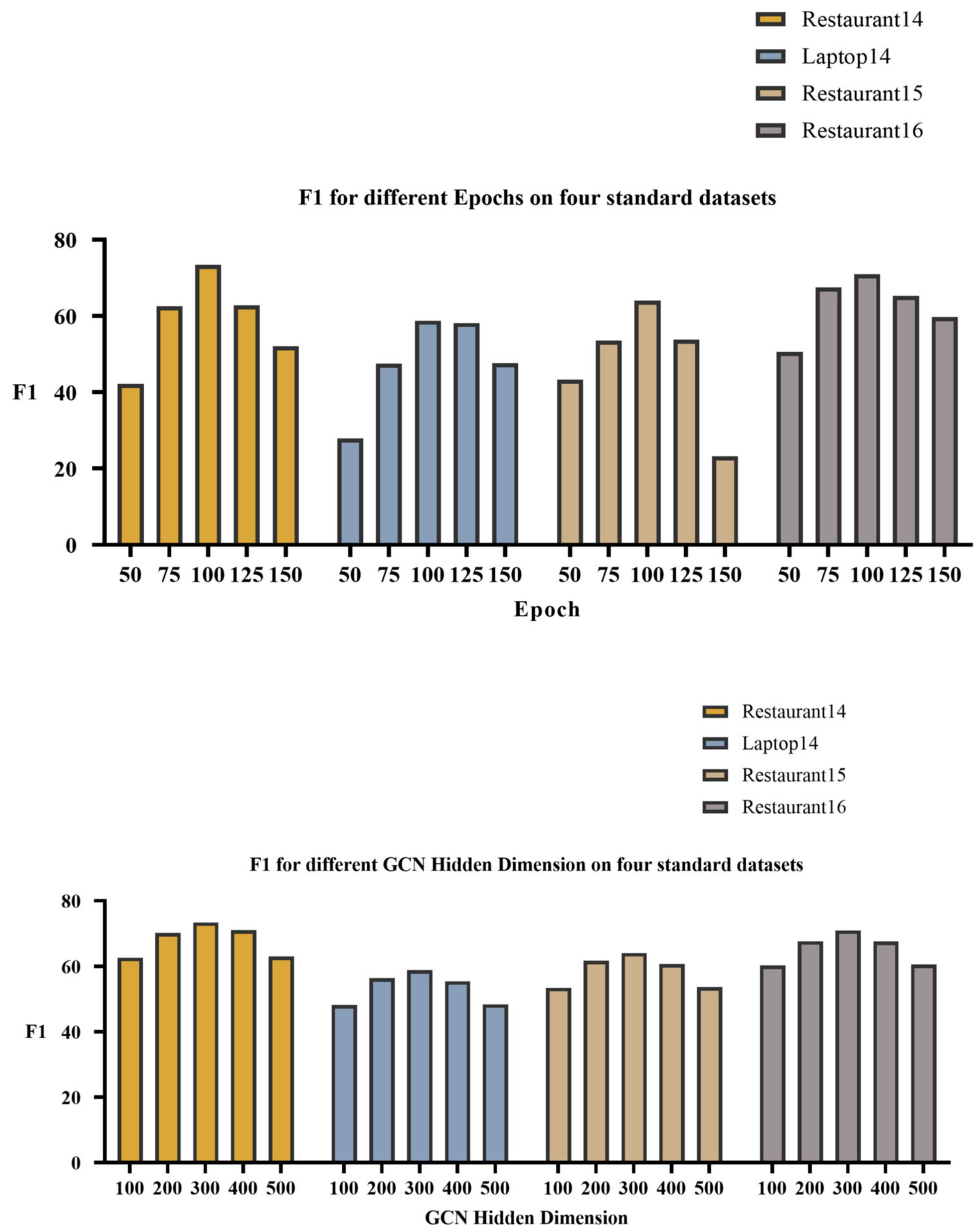
1. Introducing a modular framework that integrates BERT with Biaffine attention, significantly improving model performance.
2. Advancing sentiment analysis through graph-based extraction of linguistic features, utilizing dimensionality transformation and gating mechanisms to optimize feature relevance and filter noise.
3. Incorporating the PAIF module, which enhances the model's sensitivity to predictive information by combining BERT-derived sentence representations with enriched linguistic features, thus refining task-specific focus.
4. Empirically validating the Fpa-GCN model's superiority over state-of-the-art methods, providing compelling evidence of its efficacy.

Table 11 F1 Scores Across Training Epochs and GCN Hidden Dimensions

		F1			
		Restaurant14	Laptop14	Restaurant15	Restaurant16
Epoch	50	42.15	27.85	43.28	50.63
	75	62.50	47.50	53.50	67.50
	100	**73.38**	**58.77**	**64.02**	**70.95**
	125	62.80	58.20	53.80	65.30
	150	52.10	47.60	23.20	59.70
GCN_dim	100	62.65	48.18	53.38	60.24
	200	70.11	56.36	61.66	67.66
	300	**73.38**	**58.77**	**64.02**	**70.95**
	400	71.01	55.48	60.70	67.60
	500	62.94	48.42	53.64	60.52

GCN_dim refers to the hidden dimension of the Graph Convolutional Network (GCN)

Fig. 12 Visualization of hyperparametric sensitivity analysis



Despite its promising results, Fpa-GCN faces several challenges that warrant further investigation. For instance, the model's ability to process very long texts requires enhanced graph structure representation to capture extended dependencies effectively. Additionally, current sentiment triplet extraction methods may oversimplify sentiment categorization, failing to precisely distinguish between different types of sentiment features. Future research could extend the model to accommodate more complex tasks, such as quadruplet extraction or dialog-level quadruplet extraction, thereby broadening its applicability to more nuanced NLP scenarios.

The insights derived from the development of the Fpa-GCN model hold significant potential for advancing complex aspect-level sentiment analysis and related deep-learning tasks. Refining the feature fusion strategy and addressing the aforementioned challenges will enhance the model's capability to discern and associate intricate sentence patterns, ultimately pushing the boundaries of what is achievable in sentiment analysis and beyond.

Acknowledgements This work is supported by 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 Program of Sichuan Province (Grant no. 2023YFS0424), 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)

Author Contributions H. Y. Jiang, Conceptualization; H. Y. Jiang, Data curation; H. Y. Jiang, Formal analysis; X. L. Chen, D. Q. Miao, and X. L. Qin, Funding acquisition; X. L. Chen, Investigation; X. L. Chen, Methodology; X. L. Chen, Project administration; X. L. Qin, Resources; X. Gu, Software; X. L. Chen, Supervision; H.Y. Zhang, X. Y. Li, Validation; H. Y. Jiang, Visualization; H. Y. Jiang /Writing-original draft; X. L. Chen and P. Lu /Writing-review and editing.

Data Availability and Access Data are available for download at the following web links. https://github.com/Joeisjoejoe/Fpa-GCN

Declarations

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.

Ethical and Informed Consent for Data Used This article does not contain studies with human participants or animals. A statement of informed consent is not applicable since the manuscript does not contain any patient data.

References

1. Peng H, Xu L, Bing L et al (2020) Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In: Proceedings of the AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v34i05.6383
2. Mao Y, Shen Y, Yu C et al (2021) A joint training dual-mrc framework for aspect based sentiment analysis. In: Proceedings of the AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v35i15.17597
3. Brown TB, Mann B, Ryder N et al (2020) Language models are few-shot learners. https://doi.org/10.48550/arXiv.2005.14165. arXiv:2005.14165
4. Raffel C, Shazeer N, Roberts A et al (2019) Exploring the limits of transfer learning with a unified text-to-text transformer. https://doi.org/10.48550/arXiv.1910.10683. arXiv:1910.10683
5. Devlin J, Chang MW, Lee K et al (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies. association for computational linguistics. https://doi.org/10.18653/v1/N19-1423
6. Dozat T, Manning CD (2016) Deep biaffine attention for neural dependency parsing. https://doi.org/10.48550/arXiv.1611.01734. arXiv:1611.01734
7. Tang F, Fu L, Yao B et al (2019) Aspect based fine-grained sentiment analysis for online reviews. Inf Sci 488:190–204. https://doi.org/10.1016/j.ins.2019.02.064
8. Xiao L, Xue Y, Wang H et al (2022) Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. Neurocomputing 471:48–59. https://doi.org/10.1016/j.neucom.2021.10.091
9. Consoli S, Barbaglia L, Manzan S (2022) Fine-grained, aspect-based sentiment analysis on economic and financial lexicon. Knowl-Based Syst 247:108781. https://doi.org/10.1016/j.knosys.2022.108781
10. Phan MH, Ogunbona PO (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. https://doi.org/10.18653/v1/2020.acl-main.293
11. Wang K, Shen W, Yang Y et al (2020) Relational graph attention network for aspect-based sentiment analysis. https://doi.org/10.48550/arXiv.2004.12362. arXiv:2004.12362
12. Zhang B, Xu D, Zhang H et al (2019) Stcs lexicon: Spectral-clustering-based topic-specific chinese sentiment lexicon construction for social networks. IEEE Trans Comput Social Syst 6(6):1180–1189. https://doi.org/10.1109/TCSS.2019.2941344
13. Kiritchenko S, Zhu X, Cherry C et al (2014) Nrc-canada-2014: Detecting aspects and sentiment in customer reviews. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). https://doi.org/10.3115/v1/S14-2076
14. Dai J, Yan H, Sun T et al (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. https://doi.org/10.18653/v1/2021.naacl-main.146
15. Wang M, Cao D, Li L et al (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
16. Gui T, Zhang Q, Huang H et al (2017) Part-of-speech tagging for twitter with adversarial neural networks. In: Conference on empirical methods in natural language processing. https://doi.org/10.18653/v1/D17-1256
17. Tang H, Ji D, Li C et al (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. https://doi.org/10.18653/v1/2020.acl-main.588

18. Xiao Z, Wu J, Chen Q et al (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. https://doi.org/10.18653/v1/2021.emnlp-main.724
19. Li X, Lam W (2017) Deep multi-task learning for aspect term extraction with memory interaction. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Assoc Comput Linguistics. https://doi.org/10.18653/v1/D17-1310
20. Xu H, Liu B, Shu L et al (2018) Double embeddings and CNN-based sequence labeling for aspect extraction. In: Proceedings of the 56th annual meeting of the association for computational linguistics. association for computational linguistics. https://doi.org/10.18653/v1/P18-2094
21. Li X, Bing L, Li P et al (2018) Aspect term extraction with history attention and selective transformation. In: Proceedings of the Twenty-Seventh international joint conference on artificial intelligence, IJCAI-18. international joint conferences on artificial intelligence organization. https://doi.org/10.24963/ijcai.2018/583
22. Ma D, Li S, Wu F et al (2019) Exploring sequence-to-sequence learning in aspect term extraction. In: Proceedings of the 57th annual meeting of the association for computational linguistics. association for computational linguistics. https://doi.org/10.18653/v1/P19-1344
23. Li K, Chen C, Quan X et al (2020) Conditional augmentation for aspect term extraction via masked sequence-to-sequence generation. In: Proceedings of the 58th annual meeting of the association for computational linguistics. association for computational linguistics. https://doi.org/10.18653/v1/2020.acl-main.631
24. Wang W, Pan SJ, Dahlmeier D et al (2016) Recursive neural conditional random fields for aspect-based sentiment analysis. In: Proceedings of the 2016 conference on empirical methods in natural language processing. association for computational linguistics. https://doi.org/10.18653/v1/D16-1059
25. Wang W, Pan SJ (2018) Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. In: Proceedings of the 56th annual meeting of the association for computational linguistics. association for computational linguistics. https://doi.org/10.18653/v1/P18-1202
26. Chen Z, Qian T (2020) Relation-aware collaborative learning for unified aspect-based sentiment analysis. In: Proceedings of the 58th annual meeting of the association for computational linguistics. Association for computational linguistics. https://doi.org/10.18653/v1/2020.acl-main.340
27. He R, Lee WS, Ng HT et al (2019) An interactive multi-task learning network for end-to-end aspect-based sentiment analysis. In: Proceedings of the 57th annual meeting of the association for computational linguistics. Association for computational linguistics. https://doi.org/10.18653/v1/P19-1048
28. Wang Y, Huang M, Zhu X et al (2016) Attention-based LSTM for aspect-level sentiment classification. In: Proceedings of the 2016 conference on empirical methods in natural language processing. Association for computational linguistics. https://doi.org/10.18653/v1/D16-1058
29. Tang D, Qin B, Feng X et al (2016) Effective lstms for target-dependent sentiment classification. In: Proceedings of COLING 2016, the 26th international conference on computational linguistics: technical papers. The COLING 2016 Organizing Committee. https://aclanthology.org/C16-1311
30. Liu J, Zhang Y (2017) Attention modeling for targeted sentiment. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Assoc Comput Linguistics. https://aclanthology.org/E17-2091
31. Ma D, Li S, Zhang X et al (2017) Interactive attention networks for aspect-level sentiment classification. In: Proceedings of the Twenty-Sixth international joint conference on artificial intelligence. https://doi.org/10.24963/ijcai.2017/568
32. Tay Y, Tuan LA, Hui SC (2018) Learning to attend via word-aspect associative fusion for aspect-based sentiment analysis. In: Proceedings of the AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v32i1.12049
33. Fan Z, Wu Z, Dai XY et al (2019) Target-oriented opinion words extraction with target-fused neural sequence labeling. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1259
34. Wu Z, Ying C, Zhao F et al (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. https://doi.org/10.18653/v1/2020.findings-emnlp.234
35. Pouran Ben Veyseh A, Nouri N, Dernoncourt F et al (2020) Introducing syntactic structures into target opinion word extraction with deep learning. In: Proceedings of the 2020 conference on empirical methods in natural language processing. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.719
36. Mitchell M, Aguilar J, Wilson T et al (2013) Open domain targeted sentiment. In: Proceedings of the 2013 conference on empirical methods in natural language processing. Association for Computational Linguistics. https://aclanthology.org/D13-1171
37. Zhang M, Zhang Y, Vo DT (2015) Neural networks for open domain targeted sentiment. In: Proceedings of the 2015 conference on empirical methods in natural language processing. Association for Computational Linguistics. https://doi.org/10.18653/v1/D15-1073
38. Li X, Bing L, Li P et al (2019) A unified model for opinion target extraction and target sentiment prediction. In: Proceedings of the AAAI conference on artificial intelligence. https://doi.org/10.1609/aaai.v33i01.33016714
39. Hu M, Peng Y, Huang Z et al (2019) Open-domain targeted sentiment analysis via span-based extraction and classification. In: Proceedings of the 57th annual meeting of the association for computational linguistics. Association for Computational Linguistics. https://doi.org/10.18653/v1/P19-1051
40. Zhao H, Huang L, Zhang R et al (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. https://doi.org/10.18653/v1/2020.acl-main.296
41. Chen S, Liu J, Wang Y et al (2020) Synchronous double-channel recurrent network for aspect-opinion pair extraction. In: Annual meeting of the association for computational linguistics. https://api.semanticscholar.org/CorpusID:220047070
42. Xu L, Li H, Lu W et al (2020) Position-aware tagging for aspect sentiment triplet extraction. In: Proceedings of the 2020 conference on empirical methods in natural language processing. Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.183
43. Zhu X, Zhu L, Guo J et al (2021) Gl-gcn: Global and local dependency guided graph convolutional networks for aspect-based sentiment classification. Exp Syst Appl 186:115712. https://doi.org/10.1016/j.eswa.2021.115712
44. Feng S, Wang B, Yang Z et al (2022) Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation. Knowl-Based Syst 258:109975. https://doi.org/10.1016/j.knosys.2022.109975

45. Shi L, Han D, Han J et al (2022) Dependency graph enhanced interactive attention network for aspect sentiment triplet extraction. Neurocomputing 507:315–324. https://doi.org/10.1016/j.neucom.2022.07.067
46. Chen Z, Huang H, Liu B et al (2021) Semantic and syntactic enhanced aspect sentiment triplet extraction. https://doi.org/10.18653/v1/2021.findings-acl.128
47. Li Y, Lin Y, Lin Y et al (2022) A span-sharing joint extraction framework for harvesting aspect sentiment triplets. Knowl-Based Syst 242:108366. https://doi.org/10.1016/j.knosys.2022.108366
48. Fei H, Ren Y, Zhang Y et al (2021) Nonautoregressive encoder-decoder neural framework for end-to-end aspect-based sentiment triplet extraction. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2021.3129483
49. Pontiki M, Galanis D, Pavlopoulos J et al (2014) SemEval-2014 task 4: Aspect based sentiment analysis. In: Proceedings of the 8th international workshop on semantic evaluation (SemEval 2014). Association for Computational Linguistics. https://doi.org/10.3115/v1/S14-2004
50. Pontiki M, Galanis D, Papageorgiou H et al (2015) SemEval-2015 task 12: Aspect based sentiment analysis. In: Proceedings of the 9th international workshop on semantic evaluation (SemEval 2015). Association for Computational Linguistics. https://doi.org/10.18653/v1/S15-2082
51. Pontiki M, Galanis D, Papageorgiou H et al (2016) SemEval-2016 task 5: Aspect based sentiment analysis. In: Proceedings of the 10th international workshop on semantic evaluation. Association for Computational Linguistics. https://doi.org/10.18653/v1/S16-1002
52. Zhang C, Li Q, Song D et al (2020) A multi-task learning framework for opinion triplet extraction. https://doi.org/10.18653/v1/2020.findings-emnlp.72
53. Chen H, Zhai Z, Feng F et al (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. https://doi.org/10.18653/v1/2022.acl-long.212
54. Zhang C, Ren L, Ma F et al (2022) Structural bias for aspect sentiment triplet extraction. https://doi.org/10.48550/arXiv.2209.00820
55. Zhang C, Li Q, Song D et al (2020) A multi-task learning framework for opinion triplet extraction. In: Findings of the Association for Computational Linguistics: EMNLP 2020. Association for Computational Linguistics, new. https://doi.org/10.18653/v1/2020.findings-emnlp.72
56. Mao Y, Shen Y, Yu C et al (2021) A joint training dual-mrc framework for aspect based sentiment analysis. In: Proceedings of the AAAI conference on artificial intelligence. https://doi.org/10.48550/arXiv.2101.00816
57. Loshchilov I, Hutter F (2018) Fixing weight decay regularization in adam. https://openreview.net/forum?id=rk6qdGgCZ
58. Li B, Fei H, Li F et al (2023) DiaASQ: A benchmark of conversational aspect-based sentiment quadruple analysis. In: Findings of the association for computational linguistics: ACL 2023. Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.849
59. 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. https://doi.org/10.18653/v1/2021.acl-long.29
60. Markus Eberts AU (2020) Span-based joint entity and relation extraction with transformer pre-training. ECAI 325. https://doi.org/10.48550/arXiv.1909.07755
61. Xu L, Chia YK, 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. Association for Computational Linguistics, pp 4755–4766. https://doi.org/10.18653/v1/2021.acl-long.367
62. Zhang W, Deng Y, Li X et al (2021) Aspect sentiment quad prediction as paraphrase generation. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp 9209–9219. https://doi.org/10.18653/v1/2021.emnlp-main.726
63. Jiang H, Chen X, Miao D et al (2024) Ifusionquad: A novel framework for improved aspect-based sentiment quadruple analysis in dialogue contexts with advanced feature integration and contextual cloblock. Expert Syst Appl 261:125556. https://api.semanticscholar.org/CorpusID:273474551