Electronic
Research Archive

https://www.aimspress.com/journal/ERA

ERA, 34(5): 2897–2925.
DOI: 10.3934/era.2026132
Received: 04 November 2025
Revised: 13 January 2026
Accepted: 15 March 2026
Published: 02 April 2026

Research article

Multi-granularity features fusion with hierarchical networks for aspect-based sentiment analysis

Xiaomin Zhong*, Hengxi Di*, Xuefeng Zhao, Changze Bai and Zhaoman Zhong

School of Computer Engineering, Jiangsu Ocean University, Jiangsu 222005, China

* **Correspondence:** Email: 2001000049@jou.edu.cn, dihx@jou.edu.cn.

Abstract: Aspect-based sentiment analysis (ABSA) aims to identify and classify sentiment polarities toward specific aspects in text. To address the limitations of existing models, such as oversimplified structures and underutilized fine-grained features, we propose an ABSA model that fuses multi-granularity features with hierarchical networks. Through the construction of a multi-granularity feature extraction module, we employed a graph convolutional network (GCN) on constituent and dependency trees to capture syntactic and semantic information, augmented with external knowledge. This enables comprehensive feature extraction from four granularity levels: constituent structure, dependency relations, contextual cues, and external knowledge. To effectively fuse these diverse features, we designed a multi-layer feature fusion network (MLFF). Utilizing cross attention and orthogonal projection, the MLFF module iteratively refines feature interactions. Extensive experiments on the SemEval 2014 and Twitter datasets show that our model outperforms existing ABSA models and can effectively enhance task performance.

Keywords: aspect-based sentiment analysis; multi-granularity features; graph convolutional network; dependency tree; orthogonal projection technique

---

1. Introduction

Aspect-based sentiment analysis (ABSA) tasks are designed to assess the sentiment tendencies associated with a particular aspect or entity in a text, thereby enabling a more comprehensive understanding of the emotional information underlying the text [1]. With the onset of the big-data era, ASBA [2] has been assuming an increasingly significant role in numerous fields, including comment analysis, public opinion detection, and marketing. For instance, as depicted in Figure 1, when presented with a hotel customer's comment "The appetizers are good, but the service is too slow", the objective is to determine the sentiment polarity of the aspect terms "appetizers" and "service", where "good" and "too slow" represent positive and negative sentiment polarities, respectively. Traditional text sentiment analysis solely focuses on the overall sentiment orientation of the text. In contrast, ABSA can precisely identify the sentiment orientation of all aspects. At present, the primary challenge confronted by ABSA is how to effectively model the relationship between aspects and corresponding opinions.

In prior research, ABSA tasks have generally centered on modeling sentence semantics. These utilize neural networks and attention mechanisms to capture semantic associations among sentences, aiming to precisely determine the sentiment polarity of specific aspects [3,4]. In long sentences containing multiple aspect terms, as illustrated in Figure 1, when determining the sentiment polarity of the noun "service", previous mechanisms tended to overlook the function of the conjunction "but" and instead assign more weight to "good", resulting in an inaccurate assessment of the sentiment tendency of the aspect "service". Meanwhile, the sentiment polarity of sentences can be influenced by multiple aspects. Although traditional approaches excel in evaluating the overall sentiment of texts [2], they may face limitations in capturing multi-granularity features and global information within sentences, thereby constraining task performance to some degree.

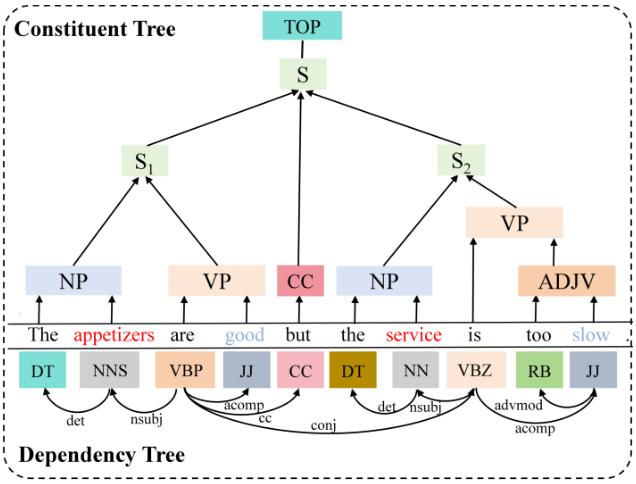

Figure 1. Example sentence with a dependency tree and a constituent tree.

To comprehensively capture diverse granular features within text sentences, researchers started to explore the integration of syntactic information, such as dependency trees [5] and constituent trees [6], into ABSA tasks. Dependency trees are capable of establishing connections among words in a sentence, whereas constituent trees can offer precise phrase segmentation and hierarchical structure, which facilitates the accurate alignment of aspect-word sentiment indicators. In Figure 1, there exists a dependency between the aspect words "appetizers" and "good", and the phrase-splitting term "but" separates "good" from "the service is too slow". This syntactic structure information provides rich context for the model, which aids in more accurately identifying the sentiment polarity of specific aspects.

In recent research, graph convolutional networks (GCN) have been extensively employed in ABSA tasks [7,8]. GCN can effectively capture semantic information within sentences by leveraging graph-structured data for information transmission, while fully considering the constituent structure and dependency relationships in sentences. Specifically, GCN can utilize structural information, such as constituent trees and dependency trees [5,6], to convert each constituent and dependency relationship in a sentence into a graph structure. Through graph convolution operations, it transfers information and integrates features, enabling a better understanding of the structural features in the sentence and enhancing the performance of sentiment analysis [9–11]. With the advancement of pre-trained models [12,13], they have demonstrated outstanding performance in diverse ABSA tasks [14,15]. The incorporation of external knowledge [16,17] has also yielded favorable results, further enhancing the accuracy of sentiment analysis.

However, although the incorporation of technologies such as constituent trees, dependency trees, and GCN has brought significant performance enhancements to ABSA tasks, existing models still encounter certain challenges. Moreover, most previous studies have demonstrated the effectiveness of single-granularity information for ABSA tasks [18,19]. However, single-granularity features are insufficient to comprehensively capture the rich information embedded in the original data. While recent multi-granularity or hybrid models have attempted to address this limitation, they still exhibit notable shortcomings. For instance, Li et al. [5] proposed the DualGCN model, which integrates semantic and syntactic features via SynGCN and SemGCN, yet it lacks mechanisms to incorporate external knowledge or to model finer-grained structural interactions. Similarly, models like BiSyn-GAT+ jointly model constituent and dependency structures but often rely on simple fusion strategies (e.g., concatenation or weighted averaging) without explicitly capturing deep synergistic relationships and orthogonal complementarity among different granularities. Furthermore, such approaches may not fully align heterogeneous features (e.g., contextual representations, external knowledge) with syntactic or constituent patterns, leading to information redundancy or conflict. Consequently, existing multi-granularity frameworks remain limited in their ability to dynamically and effectively harness the comprehensive potential of diverse feature types. Thus, how to systematically integrate and interact multi-granularity features—including syntactic, constituent, contextual, and external knowledge—through advanced fusion mechanisms remains a critical unresolved issue. Currently, the structure of ABSA models is relatively simplistic, and they fail to cohesively integrate various granular features of sentences into a single ABSA task, which constrains the model's performance to some degree.

In this paper, we propose a novel ABSA network model to tackle the aforementioned challenges. First, by constructing a multi-granularity feature extraction module, we integrate four types of text features, namely syntactic constituent structure, dependency relationship, context features, and external knowledge, to enhance the acquisition of sentiment representations in the ABSA task, thereby achieving a more comprehensive extraction of sentiment feature information. Second, an MLFF module is designed. Cross attention and orthogonal projection techniques are introduced, and repeated computational processing within the module is carried out to capture the profound and intricate interactions among features of different granularities. Finally, the processed features are input into the Softmax layer for sentiment classification. Experimental results from three publicly available datasets demonstrate that the proposed model outperforms existing state-of-the-art models. The contributions of this work can be summarized as follows:

- Proposes an ABSA model that integrates multi-granularity features fusion with hierarchical networks. The model aims to capture sentiment features at different granularities within the text and the intricate interactions among these features, thereby attaining more precise ABSA classification.
- Proposes a multi-granularity feature extraction module. Starting from four perspectives, namely syntactic constituent structure, dependency information, context features, and external knowledge,

it fully captures the complex interactions among various features in the text, enabling more accurate judgment of the sentiment polarity of aspect terms.

- Designs a multi-layer feature fusion module (MLFF). Through the cross attention and orthogonal projection technique, and after multiple repeated computations of the MLFF, mutual learning and collaboration among features of different granularities are achieved, and more refined sentiment features are obtained.
- Extensive experimental results on three benchmark datasets demonstrate that the proposed model performs better than existing models in terms of accuracy and F1-score, which proves the effectiveness of the model design.

2. Related work

2.1. Method based on a pre-training model

With the rapid advancement of deep-learning technology, pre-trained language models have been extensively employed in the domain of natural language processing (NLP), notably enhancing the performance of diverse NLP tasks. These models are typically pre-trained in an unsupervised manner on large corpora and adapted to distinct downstream tasks via fine-tuning. Among these models, bidirectional encoder representations from transformers (BERT) [12] is a pre-trained model based on the Transformer architecture. It learns text representations through unsupervised learning on a large-scale text corpus and has achieved favorable performance in numerous NLP tasks [21]. Sun et al. [22] fine-tuned the pre-trained BERT model and obtained satisfactory results on the SemEval-2014 dataset. Xu et al. [23] proposed a post-training approach based on BERT to enhance the performance during the fine-tuning phase of the ABSA task. However, a critical limitation persists: BERT primarily models sequential co-occurrence and relies on self-attention mechanisms that may not explicitly and optimally capture the explicit syntactic constraints and long-range dependency relationships crucial for aligning aspects with their specific opinion words.

RoBERTa [13] is an optimized variant of BERT that adjusts the training data and hyperparameters, utilizes a larger batch size and a longer training time, eliminates the next sentence prediction (NSP) task, and employs the dynamic masking method to enhance the generation of training data and the pre-training efficiency of the model. In comparison with the static masking of the original BERT, dynamic masking can dynamically select the words to be masked and better capture the syntactic structure and semantic relations of the language. A detailed comparison between the two is presented in Table 1.

Table 1. Comparison of static masking and dynamic masking.

Original sentence	The price is reasonable, although the service is poor.
Static masking	The price is [Mask], although the service is [Mask].
Dynamic masking	Epoch1: The price is [Mask], although the service is [Mask].
	Epoch2: The [Mask] is reasonable [Mask] the [Mask] is poor.

In a recent investigation, Liu et al. [24] proposed a novel and straightforward sentiment analysis model based on RoBERTa, which attained state-of-the-art results at that time. Shi et al. [20] obtained favorable outcomes in ABSA tasks by utilizing RoBERTa and integrating prompt representation with comparative learning. In more complex sentences, the sentiment polarity of aspect words is determined by multiple features rather than just word proximity. Therefore, RoBERTa has become a promising research direction due to its powerful semantic representation function.

2.2. *Method based on GCN*

To incorporate explicit syntactic information, GCN applied over syntactic trees has been widely explored. These approaches can be broadly categorized based on the type of syntactic formalism they leverage.

Combination of dependency trees and GCN: a prominent line of work operates on dependency parse trees, aiming to directly model relationships between aspect terms and opinion words. Models such as those by Sun et al. [25] and Zhang et al. [26] propagate information along dependency edges, while enhanced variants like RDGCN [27] refine dependency relation importance. The T-GCN model [11] further incorporates dependency type information. Despite their effectiveness, dependency-centric models often focus on binary word-to-word relations, potentially overlooking the richer hierarchical phrase-level semantics captured by constituent structures.

Combination of constituent trees and GCN: conversely, another line of research utilizes constituent (phrase-structure) trees to model sentence hierarchy. Works like Marcheggiani et al. [28] and Li et al. [29] demonstrate the utility of GCNs on constituency trees. Fei et al. [30] further incorporated heterogeneous syntactic representations into semantic role labeling (SRL) via collaborative learning of phrase boundaries during selection and semantic relationships in syntactic dependencies. Notably, Liang et al. [6] proposed BiSyn-GAT+, a dual syntactic model that jointly leverages both constituent and dependency structures via separate graph attention networks, representing a significant step toward multi-granularity synthesis.

While hybrid models like BiSyn-GAT+ [6] and DualGCN [5] have demonstrated the benefits of integrating multiple syntactic granularities, they are subject to several key limitations. Primarily, these models often rely on shallow fusion strategies (e.g., concatenation or simple attention) after parallel feature extraction, which fails to model deep, hierarchical interactions and synergistic complementarity between different feature types throughout the encoding process. Furthermore, they are predominantly knowledge-agnostic, focusing on internal sentence structures while seldom incorporating external, domain-agnostic knowledge to disambiguate context or reinforce sentiment cues. This can lead to issues of redundancy and conflict when features from multiple sources are simply aggregated without a mechanism to regulate information flow and suppress noise. To address these gaps, a more sophisticated integration framework is needed—one that not only jointly encodes diverse structural and contextual features but also aligns them through dedicated mechanisms. This insight directly motivates our proposed MLFF module, which employs cross attention and orthogonal projection to foster hierarchical interaction and integrate external knowledge within a unified model.

2.3. *Comparative analysis and novelty positioning*

While existing hybrid ABSA models have demonstrated the effectiveness of integrating multiple syntactic and semantic sources, they are largely constrained by two interrelated limitations: (1) shallow fusion strategies that rely on simple concatenation or averaging of independently extracted features, thereby failing to capture deep, hierarchical interactions, and (2) a knowledge-agnostic architecture that disregards the potential of external knowledge to resolve contextual ambiguities and strengthen sentiment signals, which may result in information redundancy or conflict during feature aggregation.

To address these shortcomings, we propose an ABSA model that integrates multi-granularity features fusion with hierarchical networks. A comparison of the proposed model with other representative models is shown in Table 2.

Table 2. Innovative comparative analysis of representative ABSA models.

Model	Signals used	Fusion stage	Interaction operator	Key innovation
DualGCN [5]	Semantic + syntactic (dependency)	Late (post-GCN)	Concatenation + classifier	First to jointly model semantic and syntactic graphs but uses simple concatenation.
BiSyn+GAT+ [6]	Syntactic (dependency + constituent)	Late (post-GAT)	Weighted averaging/concatenation	Jointly leverages dual syntactic trees, but fusion remains additive/concatenative.
KGAN [31]	Contextual + external knowledge	Intermediate (attention)	Knowledge-aware attention	Integrates knowledge graphs but does not deeply fuse with syntactic structures.
Proposed Model	Contextual + syntactic (Dep & Con) + external knowledge	Hierarchical and iterative (MLFF module)	Cross attention + orthogonal projection	1) Fuses four granularities explicitly; 2) uses iterative cross attention for deep interaction; 3) employs orthogonal projection to extract complementary, non-redundant features from each source.

3. Methodology

In this section, we will elaborate on our proposed ABSA model. The overall framework is shown in Figure 2. The definition and explanation of key symbols are shown in Table 3.

Table 3. Explanation of main symbols.

Symbol	Explanation	Symbol	Explanation
S	Input sentence	$A_{i,j}^{con(n)}$	Construct the adjacency matrix of the text constituent structure
A	Aspect word sequence	$A_{i,j}^{dep}$	Construct the adjacency matrix of the text dependency relationships
C_a	Sentiment polarity	H^{con}	Structural features of sentence components
H^R	RoBERTa encoded output	H^{dep}	Sentence dependency features
W_1, W_2	Cross-attention output features	H^{kge}	External knowledge features
$Proj(\,)$	Orthogonal projection function	F	The feature vector of feature H after MLP transformation
Z^i	The i-th layer MLFF output	R	Final fusion features

Specifically, the model mainly includes three parts: (1) a text coding module, (2) a multi-granularity feature extraction module, and (3) the MLFF. The algorithm process is as follows:

Algorithm 1: ABSA based on multi-granularity features fusion with hierarchical networks

```
Input: Sentence S={$w_1,w_2,\cdots,w_n$}, Aspect words sequence A={$A_1,A_2,\cdots,A_m$}, The corresponding
       Sentiment polarities $C_a \in \{Positive, Neutral, Negative\}$
# Phase 1: Text Coding Module
    Construct a sentence-aspect pair of the input sequence: $x$=([$CLS$]$S$[$SEP$]$A$[$SEP$])
    $H^R$=RoBERTa($x$)=[$h_1^R, h_2^R, \cdots, h_n^R$]$\in R^{n\times d}\leftarrow x$=([$CLS$]$S$[$SEP$]$A$[$SEP$]) # RoBERTa processing
# Phase 2: Multi-Granularity Feature Extraction Module
    Constituent Graphs: $A^{con} \in R^{lc\times n\times n}$ #Constituent Parser
  if $w_i$ and $w_j$ are in same $ph_u^m$
      $A_{i,j}^{con(n)}=1$
  else $A_{i,j}^{con(n)}=0$
  $H^{con}$={$h_1^{con}, h_2^{con}, \cdots, h_n^{con}$}#Generate the output hidden representation through Con GCN
    Grammar information encoding: $A^{dep} \in R^{n\times n}$ #Dependency Parser
  if $link(i,j)=1$
      $A_{i,j}^{dep}=1$
  else $A_{i,j}^{dep}=0$
  $H^{dep}$={$h_1^{dep}, h_2^{dep}, \cdots, h_n^{dep}$} #Calculated through the Dep GCN
  $H^{kge}$={$h_1^{kge}, h_2^{kge}, \cdots, h_n^{kge}$} #Knowledge Graph
    $F^{con}$ = MLP($H^{con}$) , $F^{dep}$ = MLP($H^{dep}$) , $F^{kge}$ = MLP($H^{kge}$) , $F^R$ = MLP($H^R$) #MLP processing
# Phase 3: MLFF
    Context-constitutent feature representation $W_1 \leftarrow (Q\leftarrow F^R$、$K,V\leftarrow F^{con})$ #Cross attention
    Context-dependency feature representation $W_2 \leftarrow (Q\leftarrow F^R$、$K,V\leftarrow F^{dep})$
    $\widetilde{F}^{con} \leftarrow F^{con*}=Proj(F^{con},F^R)$ # Orthogonal projection
    $\widetilde{F}^{dep} \leftarrow F^{dep*}=Proj(F^{dep},F^R)$
    $Z^i \leftarrow$ Output($W_1$、$W_2$、$\widetilde{F}^{con}$、$\widetilde{F}^{dep}$、$F^{kge}$) #Calculate the result of each MLFF module
    R $\leftarrow$ Mean($Z^1,Z^2,...,Z^6$) #Mean-pooling
  Output: P $\leftarrow$ Softmax #Final sentiment classification result
  End
```

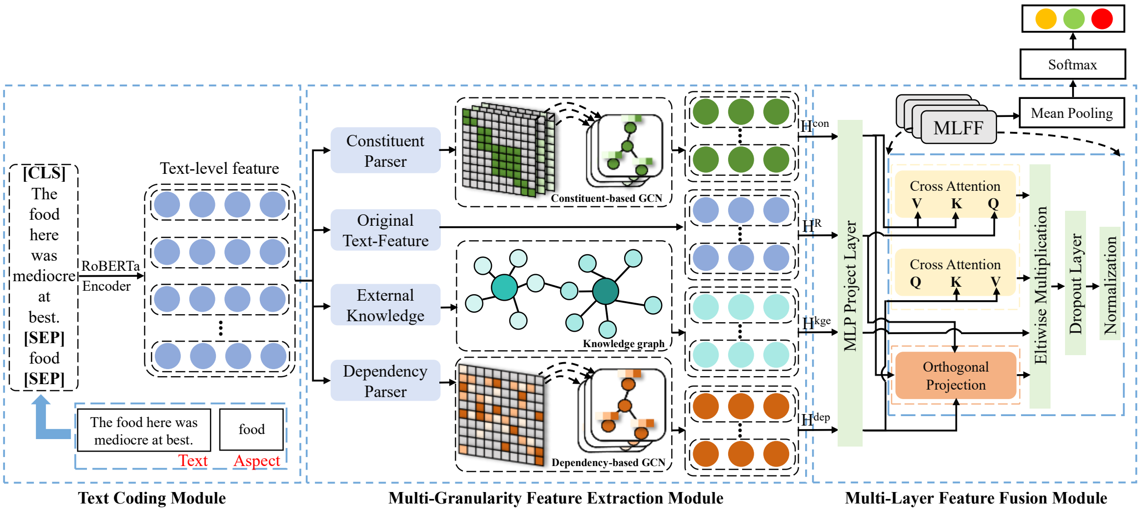

Figure 2. Overall frame structure diagram.

3.1. Text encoding module

In an ABSA task, for a given sentence consisting of n words: $S=\{w_1,w_2,\cdots,w_n\}$, the sentence consists of a sequence of aspect words of length m: $A=\{A_1,A_2,\cdots,A_m\}$, and their corresponding sentiment polarities C_a. Where w_i represents the ith word in the context, A_i denotes the ith aspect word, and aspect A is a subsequence of sentence $S,C_a\in\{Positive,Neutral,Negative\}$. To obtain a more comprehensive representation of contextual features, we employ the RoBERTa model as the word vector model to map each word into a low-dimensional real-valued vector space, constructing word vectors composed of low-dimensional real vectors. In the RoBERTa encoder, to facilitate the input of sentence and aspect word information, we construct a sentence-aspect pair as the input, which is specifically represented as $x=([CLS]S[SEP]A[SEP])$. After processing, it is represented as $H^R=\mathrm{RoBERTa}(x)=[h_1^R,h_2^R,\cdots,h_n^R]\in R^{n\times d}$, where d represents the dimension of the RoBERTa embedding, and h_i^R represents the contextual representation corresponding to the ith word.

3.2. Multi-granularity feature extraction module

Currently, the existing ABSA models only focus on extracting a single feature from the text, which lacks a relatively comprehensive multi-granularity feature extraction module. In recent years, GNNs have attracted extensive attention from researchers owing to their powerful performance in ABSA. As one of the important variants of GNNs, GCNs have achieved remarkable results when it comes to processing graph features. Additionally, the integration of external knowledge has also yielded promising outcomes in ABSA. Within this module, we design a constituent-based GCN and a dependency-based GCN to capture the semantic constituent structure and dependency relationships of sentences, respectively. We introduce external knowledge so as to assist in determining the sentiment polarity of aspect-based texts. By combining the contextual representation features output by RoBERTa, we can not only maximally retain the original features but also effectively integrate them with multi-granularity features, thereby enhancing the performance of the model.

3.2.1. Constituent-based GCN module

In the ABSA task, the word order and grammatical structure of a sentence can influence diverse sentiment expressions and the relationships between words therein. For instance, the sentiment polarity of an aspect term may be affected by the syntactic component (such as the subject, predicate, or object) in which it is located. Therefore, analyzing the word order and grammatical structure of a sentence is of crucial importance and contributes to an accurate understanding of the sentiment polarity of each aspect. To this end, leveraging the GCN, we design a constituent-based GCN (Con GCN) module. By analyzing the constituent structure of sentences, this module can more effectively capture the interactions among various aspects.

Specifically, the feature H^R output by the text encoding module is input into the constituent parser to be converted into a syntactic structure matrix (constituent tree). In the matrix, each cell represents a certain word or phrase in the sentence and describes the syntactic constituent relationships between them. Given that the change in phrase-level information between adjacent layers is minimal, and excessive alignment would waste a great deal of computational resources, for the given depth of the constituent tree, m layers are selected with an interval of 3. Moreover, the specific value of m needs to be consistent with the number of layers in the Con GCN.

Constituent graphs are specifically represented as adjacency matrices:

$$A^{con} \in R^{lc\times n\times n} \tag{3.1}$$

The specific definitions are as follows:

$$A_{i,j}^{con(n)} = \begin{cases} 1 \;\; if \; w_i \; and \; w_j \; are \, in \; same \; ph_u^m, \\ 0 \;\; otherwise \end{cases} \tag{3.2}$$

where m denotes the level of the phrase in the selected lc layer, and u denotes the constituent label associated with the phrase, e.g., S, NP, VP, etc. Subsequently, the output hidden representation is generated by Con GCN:

$$H^{con} = \{h_1^{con}, h_2^{con}, \cdots, h_n^{con}\} \tag{3.3}$$

The specific equation is as follows:

$$h_i^{con} = \sigma(\sum_{j=1}^{n} A_{ij} W^l h_j^{l-1} + b^l) \tag{3.4}$$

where W^l denotes the weight matrix, b^l represents the bias term, and σ represents the activation function.

3.2.2. Dependency-based GCN module

In specific ABSA tasks, merely considering the syntactic constituent structure of sentences is inadequate to fully capture all the information needed for sentiment classification. Sentiment expressions often involve multi-granularity dependency relationships within the lexical context. The inter-sentence dependency relationship indicates that the sentiment analysis result of one sentence may be influenced by the components of other sentences. Moreover, the aspect-based sentiment in one sentence may depend on the relevant information mentioned in another sentence. For example, when presented with the sentence “The acting in the film is excellent; however, the plot is a little disappointing”, the first clause mentions the excellent acting of the actors in the film, while the second clause refers to the film’s plot. When performing the ABSA task, the sentiment analysis result of the aspect “acting” in the first clause may be affected by the evaluation of the plot in the second clause. Therefore, in practical applications, features of sentence constituent structure and features of sentence dependencies are typically combined to improve the accuracy and performance of ABSA. To this end, we designed a dependency-based GCN (Dep GCN) module. By leveraging the dependency relationships between words in a sentence, it can extract more comprehensive semantic information. This way, we can better comprehend the structure and meaning of the sentence and enhance the model's global understanding ability.

Specifically, the Dep GCN module takes syntactic encoding as input. It encodes syntactic information using the probability matrix of all dependency arcs in the dependency parser, which is specifically represented as an adjacency matrix:

$$A^{dep} \in R^{n\times n} \tag{3.5}$$

The definition is as follows:

$$A_{i,j}^{dep} = \begin{cases} 1 & if\ link(i,j)=1, \\ 0 & otherwise \end{cases} \tag{3.6}$$

Here, *link(i,j)* indicates that there is a dependency between the *i*th and *j*th word, and the dependency graph between the words is calculated by the Dep GCN module, denoted as:

$$H^{dep} = \{h_1^{dep}, h_2^{dep}, \cdots, h_n^{dep}\} \tag{3.7}$$

The formula is as follows:

$$h_i^{dep} = \sigma(\sum_{j=1}^{n} A_{ij} W^{l} h_j^{l-1} + b^{l}) \tag{3.8}$$

where W^l denotes the weight matrix, b^l denotes the bias term, and σ is the activation function.

3.2.3. External knowledge module

Recently, in the context of addressing ABSA tasks, given that external knowledge enables the model to comprehend natural language more comprehensively, the integration of external knowledge has gained increasing popularity. Current approaches employ intricate and inefficient means to incorporate external knowledge, for example, by directly searching graph nodes. The complementarity between external knowledge and linguistic information remains to be fully explored. Zhong et al. [31] put forward a knowledge graph augmented network (KGAN) that captures sentiment feature representations from diverse perspectives, jointly integrates contextual information and knowledge-based information, and further acquires aspect-specific knowledge representations via the attention mechanism to attain more comprehensive feature representations. Consequently, we directly adopt the knowledge graphs proposed by them and, by combining the RoBERTa encoded feature vectors H^R, generate the external knowledge vector H^{kge}, as shown in Figure 3.

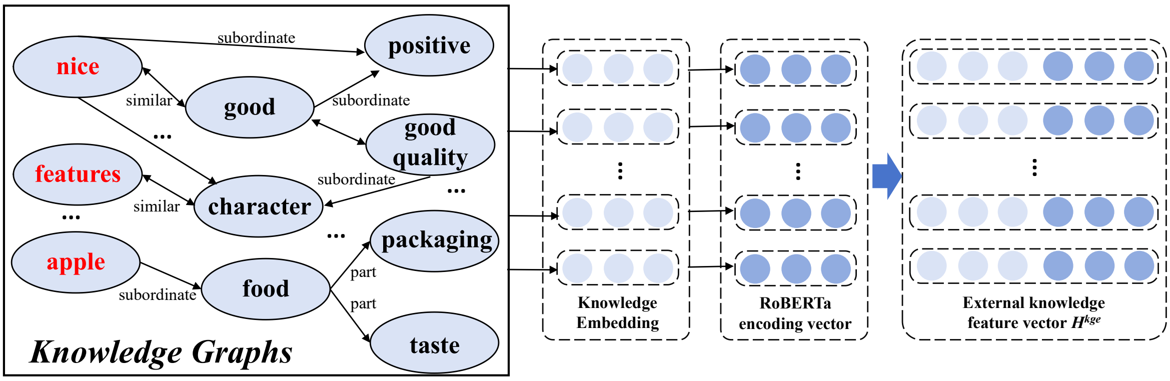

Figure 3. External knowledge structure diagram.

$$H^{kge} = \{h_1^{kge}, h_2^{kge}, \ldots, h_n^{kge}\} \tag{3.9}$$

3.2.4. Context feature module

In the text encoding module, the features generated by RoBERTa are separately fed into the constituent-based GCN, dependency-based GCN, and external knowledge modules for processing. This operation can produce more comprehensive constituent structure features and dependency relationships. However, to fully utilize the language representations learned by the pre-trained model, maximally preserve feature representations of the original context, and improve the performance of the sentiment analysis task, we directly retain features processed by RoBERTa.

Specifically, this is denoted as:

$$H^{R} = RoBERTa(x) = [h_1^R, h_2^R, h_3^R, \cdots, h_n^R] \in R^{n \times d} \tag{3.10}$$

3.2.5. MLP project layer

In ABSA tasks, the extracted features encompass sentiment information from different aspects and usually have different dimensions or feature expression modes. Especially when there is a wide variety of features, it becomes particularly crucial to handle these features more effectively. Therefore, for the four extracted features H^{con}, H^{dep}, H^{kge}, and H^{R}, an MLP module is introduced to conduct in-depth mapping and transformation of the features, so as to better capture the relationships between features and enhance the model's ability to represent features.

In the hidden layer, transformation is carried out through a linear layer and the ReLU function, followed by a dropout operation. During the training process, some elements in the input vector are randomly set to 0 to prevent overfitting. Finally, in the output layer, the output vector of the hidden layer undergoes another linear transformation to obtain the final feature vector outputs, which are specifically represented as F^{con}, F^{dep}, F^{kge}, and F^{R}.

3.3. MLFF

Specifically, four granularity features F^{con}, F^{dep}, F^{kge}, and F^{R} are outputted, which describe the feature information of the text from different perspectives. At present, the methods for processing and fusing multiple granularity features are relatively simplistic, failing to fully leverage their respective characteristics and relevant advantages, which easily leads to a relatively high data complexity. Therefore, inspired by [32,33], we design an MLFF: we use cross attention and orthogonal projection techniques to extract more representative features. After multiple repeated processes of MLFF, the redundancy of features is reduced to a certain extent, and thus the computational cost is decreased. Finally, we perform feature fusion and determine the sentiment polarity. Specifically, it includes three modules: (1) a cross-attention module, (2) an orthogonal projection module, and (3) a feature fusion module. The calculation process of each layer of MLFF is shown in Algorithm 2.

We present the parameter counts and computational overhead of the key components of our proposed model based on the RoBERTa-based architecture. The multi-granularity feature extraction and MLFF modules introduce approximately 4.2 million parameters, amounting to a 7.5% increase over the base RoBERTa model, which contains approximately 110 million parameters. A detailed breakdown reveals that each GCN module contributes 0.4 million parameters, the external knowledge integration module adds 0.8 million parameters, and the core MLFF module, comprising MLP projectors, cross-attention mechanisms, and fusion layers, accounts for 2.6 million parameters.

```
Algorithm 2: One layer of MLFF
Input:
    $F^R$: Contextual feature from RoBERTa, shape: (batch_size, seq_len, d_model), where d_model = 768.
    $F^{con}$: Constituent structure feature from Con GCN, shape: (batch_size, seq_len, d_con). d_con is set to 300.
    $F^{dep}$: Dependency relation feature from Dep GCN, shape: (batch_size, seq_len, d_dep). d_dep is set to 300.
    $F^{kge}$: External knowledge feature, shape: (batch_size, seq_len, d_kge). d_kge is set to 300.
Procedure:
    Cross-Attention: We employ standard multi-head Transformer cross-attention (8 heads).
        W1 = CrossAttention(query=$F^R$, key=$F^{con}$, value=$F^{con}$)
        W2 = CrossAttention(query=$F^R$, key=$F^{dep}$, value=$F^{dep}$)
    Orthogonal Projection: As per Eqs (13)–(16), compute $F^{con^*}$ and $F^{dep^*}$ from $F^{con}$ and $F^{dep}$ with
        respect to $F^R$.
  • EMFH-style Fusion: We adapt the core idea of scalable higher-order pooling from EMFH. The five features $W_1$,
    $W_2$, $\widetilde{F}^{con^*}$, $\widetilde{F}^{dep^*}$, $F^{kge}$ are first concatenated along the feature dimension. Then, an element-wise multiplication
    layer followed by a Dropout (p=0.1) and a LayerNorm operation is applied to capture their inter-feature correlations,
    producing a fused representation $Z^i$.
Output: The fused representation $Z^i$ for layer $i$, shape: (batch_size, seq_len, d_fusion).
End.
```

3.3.1. Cross-attention module

The attention mechanism has been widely applied in sentiment analysis tasks [34]. In particular, cross attention can effectively capture the complex associations between texts and better take into account the collaborative interaction among various granularity features, thereby improving the performance of the ABSA task. Specifically, for the multi-granularity features F^{con}, F^{dep}, F^{kge}, and F^R, considering that the F^{kge} feature already contains rich semantic and syntactic information, we do not perform any processing on the F^{kge} feature and directly pass it to the feature fusion module. For the three features F^{con}, F^{dep}, and F^R, we design a two-layer cross attention to process the F^{con}, F^R, and F^{dep}, F^R features, respectively. We use the attention mechanism to combine the two interactively and obtain the context-constituent and context-dependency feature representations.

After obtaining features F^{con} and F^R, we use F^R as the query Q of the attention layer and F^{con} as the key K and value V. Then, through calculation, we get the context-constituent feature representation W_1:

$$W_1 = softmax(\frac{Q_1 K_1^T}{\sqrt{d_{k_1}}})V_1 \tag{3.11}$$

Similarly, for the input features F^{dep} and F^R, we take F^R as the query Q of the attention layer and F^{dep} as the key K and value V. Then, through calculation, we obtain the context-dependency feature representation W_2:

$$W_2 = softmax(\frac{Q_2 K_2^T}{\sqrt{d_{k_2}}})V_2 \tag{3.12}$$

where d_k denotes the vector dimension of the kth header.

3.3.2. Orthogonal projection module

In view of the diversity and complexity of multi-granularity features, to some extent, this increases the computational cost of the model and is prone to data redundancy. To address this issue, the orthogonal projection technique [35] is innovatively introduced. By performing a two-dimensional transformation on the feature vector space, multiple benefits are achieved.

Specifically, as shown in Figure 4, the sentence structure feature F^{con} is first projected onto the context feature F^R as follows:

$$F^{con^*} = Proj(F^{con}, F^R) = \frac{F^{con} \cdot F^R}{|F^R|} \frac{F^R}{|F^R|} \tag{3.13}$$

where $Proj(\ )$ is the projection function.

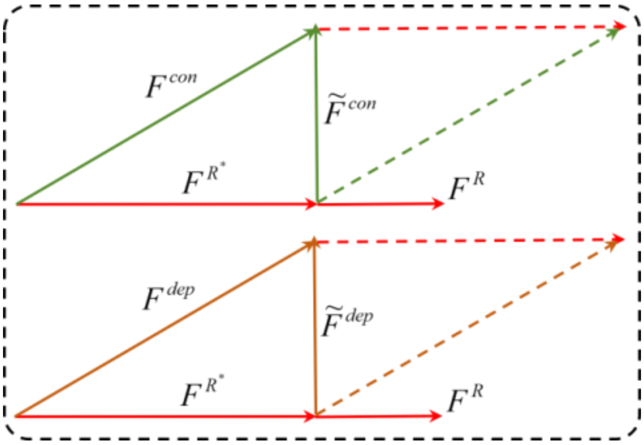

Figure 4. Structure diagram of orthogonal projection technique.

Second, project in the orthogonal direction of feature F^{con} to obtain a more refined feature, as shown below:

$$\widetilde{F}^{con} = Proj(F^{con}, (F^{con} - F^{con^*})) \tag{3.14}$$

Similarly, the dependency relation feature F^{dep} is projected onto F^R, as shown below:

$$F^{dep^*} = Proj(F^{dep}, F^R) = \frac{F^{dep} \cdot F^R}{|F^R|} \frac{F^R}{|F^R|} \tag{3.15}$$

$$\widetilde{F}^{dep} = Proj(F^{dep}, (F^{dep} - F^{dep^*})) \tag{3.16}$$

Through the orthogonal projection operation, not only is the dimensionality of the features reduced, but also the alignment ability between local features and global context features is enhanced, which significantly improves the model's performance.

The orthogonal projection operation is introduced to enforce a specific inductive bias crucial for ABSA: extracting aspect-specific syntactic signals that are complementary, rather than redundant, to the general contextual representation. Geometrically, projecting the syntactic feature F^{con} onto the contextual feature F^R isolates the component of syntax that is linearly aligned with the general sentence semantics. The residual vector $\widetilde{F}^{con}$, which lies in the subspace orthogonal to F^R, represents syntactic information that is unique and not explainable by the pre-trained language model's context alone. In ABSA, this orthogonal component often captures precise, aspect-opinion structural alignments (e.g., a dependency link negated by a distant contrastive word like "but") that are vital for correct classification but may be diluted in the holistic contextual embedding. Similarly, processing F^{dep} ensures that dependency relations contribute non-redundant information. This mechanism actively reduces feature redundancy and promotes the fusion of diverse, complementary clues.

3.3.3. Feature fusion module

In previous research work, most of the designed models only integrate features of 1–2 granularities. If one wants to integrate features of more granularities, it will increase the complexity and computational cost of the model and may even lead to significant data redundancy. Based on the four types of features output by the cross attention and orthogonal projection modules, inspired by the scalable co-attention fusion method MAMN for multiple features proposed by Xue et al. [36], the scalable multimodal decomposed higher-order pooling mechanism (EMFH) they designed is adopted. This mechanism effectively expands and mines the relevant information among multi-granularity features and has the ability to extract complex interactions between multi-granularity attention maps, greatly enhancing the scalability and efficiency of feature fusion.

Specifically, for the five input features W_1, W_2, $\widetilde{F}^{con}$, $\widetilde{F}^{dep}$, and F^{kge}, first, the element-wise multiplication layer is used to integrate the correlations among multiple features. A dropout layer is utilized to prevent overfitting. Subsequently, the output is normalized to accelerate training and enhance the model's generalization ability. Considering that a single fusion interaction is insufficient to fully exploit the comprehensive potential among multi-granularity features, the MLFF is calculated repeatedly multiple times. Finally, mean-pooling is performed to compute the average feature output, and a feature R with four different feature fusions is obtained, as shown below:

$$Z^i = Norm(\widetilde{U}_{w_1}^T W_1 \circ \widetilde{U}_{w_2}^T W_2 \circ \widetilde{U}_{con}^T \widetilde{F}^{con} \circ \widetilde{U}_{dep}^T \widetilde{F}^{dep} \circ \widetilde{U}_{kge}^T F^{kge}) \quad (3.17)$$

$$R = Mean(Z^1, Z^2, ..., Z^m) \quad (3.18)$$

where all $\widetilde{U}$ denote the learnable weight parameters, *Norm* denotes the normalization layer, Z^i denotes the output within the module at layer i, R denotes the output of the final feature, m denotes the number of module layers, and *Mean*( ) denotes the average pooling function.

Finally, leveraging the final output feature R obtained from MLFF, the final prediction probabilities for sentiment classification are derived through the Softmax function, as shown below:

$$P = Softmax(W_R R + b_R) \quad (3.19)$$

where W_R is the weight matrix, and b_R is the bias term.

4. Experiment

4.1. Datasets

To validate the effectiveness of the proposed model, a series of comprehensive experiments was carried out on three benchmark datasets. Among these datasets, the Laptop and Restaurant datasets are sourced from SemEval 2014 [37], and the Twitter dataset was gathered from Twitter by Dong et al. [38]. The detailed information is presented in Table 4.

Table 4. Dataset-specific information.

Dataset	Positive		Neutral		Negative		Total
	Train	Test	Train	Test	Train	Test	
Laptop	976	337	455	167	851	128	2944
Restaurant	2164	727	637	169	807	196	5700
Twitter	1507	172	3016	336	1528	169	6728

4.2. Evaluation index

In this paper, we use accuracy (Acc) and F1 value (F1) as evaluation metrics to assess our model, as shown in Eqs (20)–(23).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.1}$$

$$P = \frac{TP}{TP + FP} \tag{4.2}$$

$$R = \frac{TP}{TP + FN} \tag{4.3}$$

$$F1 = \frac{2 \cdot P \cdot R}{P + R} \tag{4.4}$$

where TP is a positive sample with a positive prediction, TN is a negative sample with a negative prediction, FP is a negative sample with a positive prediction, and FN is a positive sample with a negative prediction.

4.3. Hyperparameter settings

In the experiments, we implemented our model using RoBERTa-base as the backbone, which consists of 12 layers with 768 hidden units. The dropout rate was set to 0.3, and SuPar5 was employed as the parser to acquire the constituent tree and dependency tree. Based on the results of the specific experiment and the model's performance, the learning rate was set to 2×10^{-5}, the number of cross-attention heads was set to 8, the training batch size was set to 16, the epochs for each model were 15, and the Adam optimizer with a learning rate of 2×10^{-5} was utilized. For the three distinct types of experimental datasets, the number of layers of Con GCN and Dep GCN was set to (6, 3), (3, 3), and (2, 3), respectively, and the β coefficients were set to (0.12, 0.12, 0.07). A 3-layer l_c was selected as the constituent tree, and its performance was optimized. The MLFF consists of 6 layers, and the hyperparameter margin was set to 0.2. Specifically, experiments on each dataset were carried out multiple times, and the average values were calculated to avoid random results. The following comparison models are all based on the same experimental environment.

4.4. Baseline methodology

To verify the feasibility of our model in the ABSA task, we compared it with baseline models based on three different language encoders on three public datasets. The selected encoders include LSTM, BERT, and RoBERTa. The specific details are as follows.

4.4.1. Based on the LSTM model

(1) R-GAT [39]: a novel aspect-oriented dependency tree structure is defined, and the new dependency tree is encoded by the proposed graph attention network for emotion classification.

(2) DGEDT [40]: a dependency graph-enhanced dual transform network is proposed, which takes into account both learning from converters to planar representation and learning from corresponding dependency graphs to graph-based representation in an iterative manner.

(3) DualGCN [5]: a two-graph convolutional network model is proposed based on the complementarity of syntactic structure and semantic association.

(4) Hete-GNNs [41]: the sentence syntax tree, word relations, and viewpoint dictionary information are encoded in a unified framework for emotion classification.

4.4.2. Based on the BERT model

(1) R-GAT+BERT [39]: based on the attention network of the proposed relationship graph, the BERT pre-training model is combined.

(2) DGEDT+BERT [40]: the dependency graph-enhanced double transform network is combined with the pre-trained model BERT.

(3) BERT4GCN [42]: syntactic sequence features from the BERT model and PLM and syntactic knowledge from dependency graphs are integrated.

(4) AG-VSG+BERT [43]: includes attention-based auxiliary schema representation and variable sentence representation.

(5) RSSG+BERT [44]: the convolutional layer is used to comprehensively consider the relationships between word features in the local domain to reduce problems caused by incorrect syntactic dependencies.

(6) KE-IGCN [45]: this paper proposes a knowledge interaction mechanism, introduces a seed graph construction strategy, and designs a knowledge-enhanced interaction graph convolutional network.

(7) IA-HiNET-BERT [46]: in this paper, a hierarchical network oriented to inter-aspect modeling is proposed to further realize the sentence-level emotion analysis character based on aspect words.

(8) SSEMGAT-BERT [47]: constituent trees and aspect-aware attention are introduced to assign attention weights for specific aspects between contexts to enhance syntactic and semantic features.

4.4.3. Based on the RoBERTa model

(1) RGAT-RoBERTa [48]: the diagram attention network is combined with the features learned by RoBERTa, and the RoBERTa model is fine-tuned.

(2) RoBERTa4GCN [42]: based on the characteristic information learned by RoBERTa, it is fused with different GCNS.

(3) RoBERTa-MLP [42]: the feature information learned by RoBERTa is combined with the multi-layer perceptron.

(4) DGEDT-RoBERTa [49]: the proposed dependency graph-enhanced dual-transform network is combined with the pre-trained model RoBERTa.

(5) HRLN-RoBERTa [49]: a new heterogeneous reinforcement learning network based on RoBERTa is proposed.

(6) PRoGCN [50]: based on RoBERTa, a fusion heterogeneous feature representation model is proposed, and cue learning is introduced.

(7) PRCL-GCN [51]: task-specific prompt templates are designed to guide the fine-tuning process of PLM in ABSA tasks, and double affine attention mechanism is used to further extract basic emotional features from prompt representations.

4.5. Experimental results and model analysis

In this paper, comparative experiments are conducted with 18 groups of models on three datasets. The data of the comparative models are sourced from the references. The specific experimental results are presented in Table 5.

In the LSTM group, when comparing four different baseline models, the proposed model shows an average improvement of nearly 3–4% in both Acc and F1 over the best-performing comparative models on the three datasets. Even with the same LSTM encoder, significant performance enhancements can be achieved, indirectly validating the rationality and effectiveness of the network structure designs in this paper. Second, compared with the current BERT-based ABSA models, our proposed model significantly outperforms the selected BERT-based methods in terms of Acc and F1 scores. Specifically, on the Laptop dataset, compared with the KE-IGCN model, the proposed model improves the Acc and F1 scores by 1.09% and 1.56%, respectively. Additionally, compared with the RSSG+BERT model on the Restaurant dataset, our model enhances the Acc by 1.05%. Moreover, it also shows a notable lead on the Twitter dataset. Finally, even when using the same BERT encoder, our model still achieves certain performance improvements. RoBERTa is an optimized version of the BERT model. Building on the RoBERTa model, our model fully leverages the respective advantages of GCN, RoBERTa, and knowledge graph to achieve multi-aspect extraction of text features. Once optimized by MLFF, it can perform well in the sentiment classification task. Among the comparative models based on RoBERTa, the proposed model outperforms the comparative models in overall performance on the three datasets. Although the Acc on the Laptop dataset is slightly lower than that of PRC-GCN, it does not affect its good overall performance. Moreover, the use of RoBERTa can effectively streamline the calculations, improve the operation efficiency, and fully utilize RoBERTa's ability in modeling context representation.

In conclusion, the favorable results obtained from the experiments prove the success of our model in the ABSA task. This also indicates that the model can extract effective multi-granularity text feature information to achieve more accurate sentiment classification.

Table 5. Comparative experimental results (%).

Type	Models	Laptop		Restaurant		Twitter	
		Acc	F1	Acc	F1	Acc	F1
LSTM	R-GAT	77.42	73.76	83.30	76.08	75.57	73.82
	DGEDT	76.80	72.30	83.90	75.10	74.80	73.40
	DualGCN	78.48	74.74	84.27	78.08	75.92	74.29
	Hete-CNNs	74.08	69.45	81.91	73.74	72.80	71.36
BERT	R-GAT+BERT	78.21	74.07	86.60	81.35	76.15	74.88
	DGEDT+BERT	79.80	75.60	86.30	80.00	77.90	75.40
	BERT4GCN	77.49	73.01	84.75	77.11	74.73	73.76
	AG-VSG+BERT	79.92	75.85	86.34	80.88	76.45	75.04
	RSSG+BERT	80.30	76.80	87.00	81.30	77.00	76.00
	KE-IGCN	81.06	77.89	86.70	81.05	-	-
	IA-HiNET-BERT	79.45	76.57	86.79	81.84	75.88	75.36
	SSEMGAT-BERT	80.06	76.78	86.42	79.72	76.81	76.10
RoBERTa	RGAT-RoBERTa	77.43	74.21	82.76	75.25	75.43	74.04
	RoBERTa4GCN	81.80	78.16	86.23	78.61	74.75	74.00
	RoBERTa-MLP	81.11	77.11	86.79	79.76	72.76	71.73
	DGEDT-RoBERTa	60.80	48.60	66.80	27.40	61.40	57.90
	HRLN-RoBERTa	66.20	58.40	66.90	35.90	68.90	67.00
	PRoGCN	81.82	78.74	87.32	81.32	76.45	74.82
	PRCL-GCN	**82.29**	79.37	87.86	81.64	77.46	76.35
Proposed model-LSTM		80.52	76.17	86.21	79.67	76.09	74.82
Proposed model-BERT		81.17	78.23	87.71	82.34	77.84	76.11
Proposed model-RoBERTa		82.15	**79.45**	**88.05**	**82.76**	**78.03**	**76.78**

Note: "-" means not reported. The best result is in bold, and the suboptimal result is underlined.

4.6. Ablation experiment

To further investigate the rationality and advantages of each component in our model, we conducted ablation studies by systematically decomposing the model. Specifically, we removed RoBERTa (replaced by BERT), constituent-based GCN, dependency-based GCN, external knowledge, orthogonal projection, and cross attention to verify the effective contribution of the removed parts to the overall model. Meanwhile, to control the influence of parameters on the experiment, the hyperparameters for each group of ablation experiments were set to be the same. The specific experimental results are shown in Table 6 ("w/o" represents "without").

Table 6. Ablation study (%).

Model	**Laptop**		**Restaurant**		**Twitter**	
	Acc	**F1**	**Acc**	**F1**	**Acc**	**F1**
w/o RoBERTa	81.17	78.23	87.71	82.34	77.84	76.11
w/o constituent-based GCN	79.37	75.86	84.07	77.47	74.40	73.59
w/o dependency-based GCN	80.63	77.22	86.24	80.03	75.01	74.25
w/o external knowledge	81.11	77.68	86.79	81.41	76.45	74.91
w/o all syntax (only KG)	80.05	76.80	85.20	79.45	75.80	74.60
w/o orthogonal projection	80.16	76.80	87.24	80.78	74.70	73.85
w/o cross attention	78.95	76.34	83.80	75.91	75.23	72.68
Proposed model-RoBERTa	**82.15**	**79.45**	**88.05**	**82.76**	**78.03**	**76.78**

In conclusion, across the three public datasets, each component of the model presented in this paper significantly enhances the model's performance. Under the same experimental conditions, word vectors trained with RoBERTa exhibit an approximately 1% improvement in all performance aspects when processing the datasets compared to those trained with BERT. It is evident that RoBERTa performs effectively in text word-vector training. When external knowledge is removed, extracting sentiment features solely via constituent-based GCN and dependency-based GCN can still produce favorable results. However, the incorporation of external knowledge can offer the most comprehensive and rich entity information of aspect terms, aiding the model in more accurately comprehending the semantic background and associated entities behind them. When the constituent-based GCN or dependency-based GCN module is removed, the syntactic constituent structure information and dependency relations of the text are overlooked; this leads to a significant decline in the Acc and F1 values, and the overall performance demonstrates certain shortcomings. In the feature fusion module, when orthogonal projection and cross attention are removed, respectively, the model loses a certain level of accuracy and expressive capacity when handling complex sentence structures and semantic relations, thus having the most substantial impact on the model performance. Particularly, when cross attention is removed, the overall performance declines by 2%–4%, which also suggests that MLFF has a positive influence on our model.

To address whether the performance gain stems merely from increased model capacity rather than the specific interactive design of MLFF, we compare the parameter count of our MLFF module with a capacity-matched alternative. We construct a control module by stacking standard transformer blocks (with comparable hidden dimensions and layer numbers) that process the concatenated multi-granularity features. This control module has approximately 2.8 M parameters, which is slightly higher than our MLFF module (2.6 M). When replacing MLFF with this Transformer-block stack, the accuracy on the Restaurant test set drops by 1.3% (from 88.05% to 86.75%). This indicates that the performance improvement is attributable to the dedicated cross-attention and orthogonal projection mechanisms within MLFF, rather than simply adding more parameters.

To better verify the rationality of the orthogonal projection technique adopted in the MLFF module, an ablation experiment was conducted solely on the orthogonal projection technique. By comparing it with the simplified baseline method, the specific experimental results are shown in Table 7.

Table 7. Orthogonal projection ablation experiment (%).

Type	Laptop		Restaurant		Twitter	
	Acc	F1	Acc	F1	Acc	F1
Gating	81.65	78.70	87.42	81.95	77.45	76.20
LayerNorm re-weighting	81.40	78.30	87.18	81.60	77.20	75.95
Residual MLP	81.85	78.95	87.60	82.10	77.60	76.30
Orthogonal projection (ours)	**82.15**	**79.45**	**88.05**	**82.76**	**78.03**	**76.78**

As shown in the table, on the three public datasets, the MLFF module has achieved the best results by using the orthogonal projection technique. Although the residual MLP has also achieved good results, the use of the orthogonal projection technique can effectively reduce redundancy and extract the complementary syntactic context structure information, which is crucial for fine-grained ABSA tasks. These results verify the rationality of the design using orthogonal projection technology.

4.7. Effect of different hyperparameters

The selection of hyperparameters has a significant impact on the performance of the overall model. Therefore, we conducted extensive experiments taking into account the number of cross-attention heads and the number of GCN and MLFF layers to demonstrate the rationality of each hyperparameter design of the model.

4.7.1. Influence of different numbers of cross-attention heads

According to previous studies, cross attention plays a crucial role in the interaction of various types of features within the modality. To a certain extent, the number of its heads determines the model's ability and efficiency in handling different information interactions. To this end, we selected the number of heads as 2, 4, 8, 12, and 16 for validation in a specific experiment to obtain the performance of the number of attention heads on the datasets, as shown in Figure 5.

4.7.2. Influence of different GCN layers

In previous studies, the number of GCN layers has always been regarded as an important parameter affecting model performance. To find the optimal number of GCN layers, we set the number of layers from 1 to 9 and conducted experiments on three datasets to explore the relationship between the number of GCN layers and the Acc and F1 values.

The experimental results are shown in Figure 6. When the overall number of GCN layers is 1, an insufficient number of GCN layers fails to fully utilize feature representations of text interactions. In contrast, for the Laptop and Restaurant datasets, when the number of Dep GCN and Con GCN layers is 3, the model can achieve optimal performance. However, the model's performance on the Twitter dataset is suboptimal. Considering that the Twitter dataset is relatively large in terms of both quantity and data variety, excessive increases in the number of GCN layers add a large number of parameters, leading to a lack of overall feature-learning ability in the model. Therefore, for the Twitter dataset, setting the number of Dep GCN and Con GCN layers to 2 enables the model to achieve optimal performance, and the final experimental results validate this hypothesis.

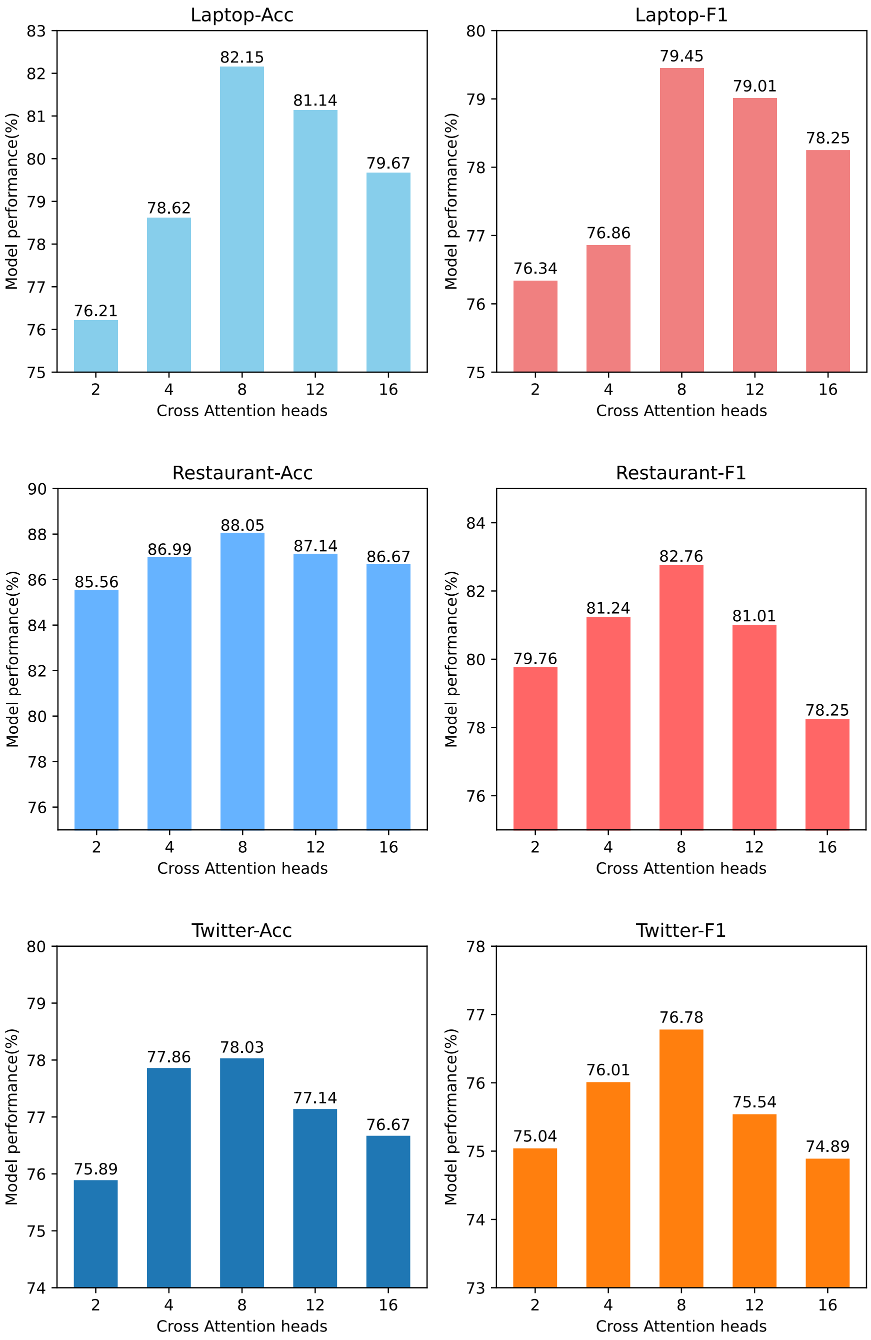

Figure 5. Influence of different numbers of cross-attention heads.

4.7.3. Influence of the number of MLFF layers

Given that the model extracts multi-granularity features from aspect-based text, the information generated has a relatively rich structure and complex emotional features. Therefore, the MLFF design is particularly important in the in-depth filtering and processing of feature information so as to reduce data redundancy and streamline emotional features. Consequently, the determination of the number of MLFF layers has a crucial impact on the overall performance. To determine the optimal number of layers, we varied the number of layers from 1 to 9 and conducted experiments on three datasets to investigate the relationship between the number of layers and the accuracy and F1 scores.

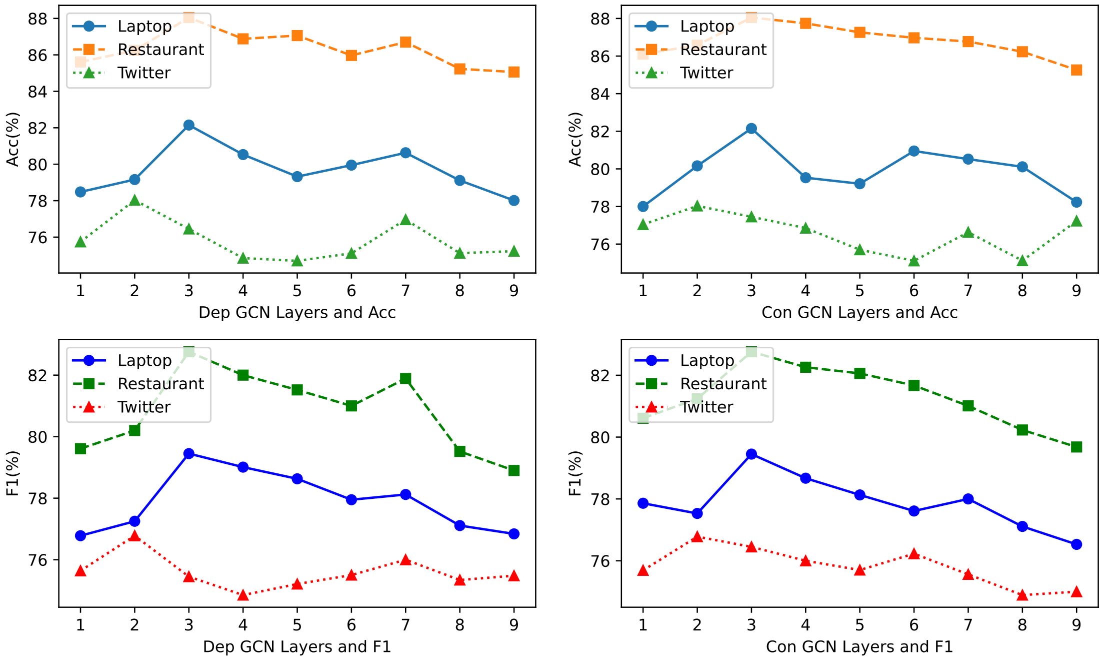

Figure 6. Influence of different GCN layers.

As shown in Figure 7, on the three public datasets, the number of layers of the MLFF module has a great impact on the model performance, regarding both Acc and F1. As the number of layers increases, the model performance also rises, and the overall model performance is best at 6 layers. On the contrary, the overall performance is poor, so 6 layers are finally selected as the optimum MLFF layer number.

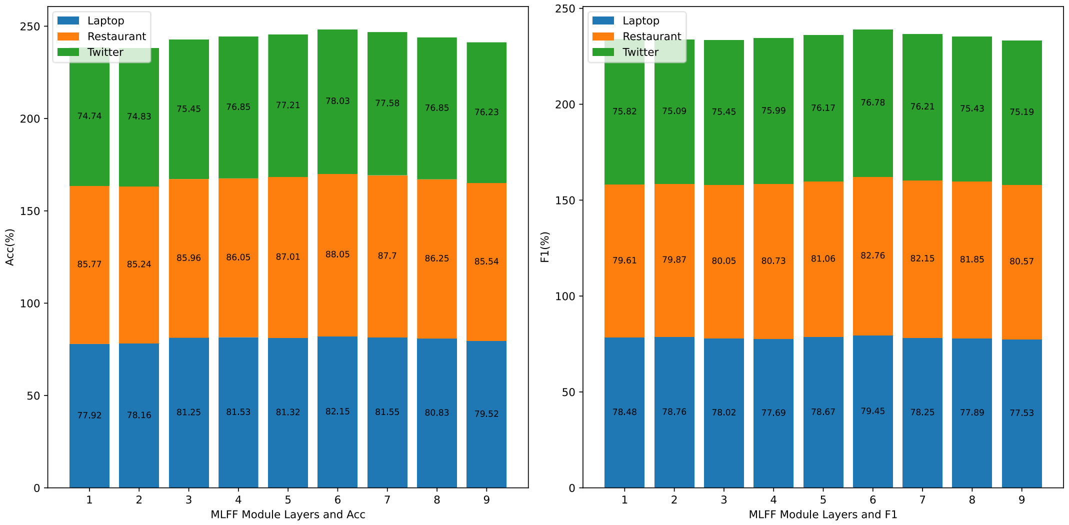

Figure 7. Influence of the number of MLFF layers.

4.8. Case studies and error analysis

To evaluate the performance and manifestation of our model in practical applications, this paper selects individual cases from three datasets and compares them with the DGEDT, RGAT, and DualGCN models in Table 4. Detailed experimental results are shown in Table 8.

Table 8. Case studies.

Sentence	RGAT	DGEDT	DualGCN	Ours	Label
1. The **food** here was mediocre at best.	Neg√	Neg√	Neg√	Neg√	Neg.
2. Great **food** but the **service** was dreadful!	Pos√,Neu×	Pos√,Pos×	Pos√,Neu×	Pos√,Neg√	Pos,Neg.
3. Skip **dessert**.	Neu×	Neu×	Neu×	Neg√	Neg.
4. I drank a **Bombay** that was big enough for two.	Pos√	Pos×	Pos√	Pos√	Pos.
5. I never tried any external **mics** with that iMac.	Neg×	Neg×	Neu√	Neu√	Neu.

As shown in Table 8, "Pos.", "Neg.", and "Neu." represent positive, negative, and neutral sentiment orientations, respectively. In the examples, the aspect words in the sentences are bolded and underlined. Here, "√" indicates that the model's prediction is correct, and "×" indicates that the prediction is wrong. The column name "Label" represents the true sentiment polarity of the aspect words. Sentence 1 is relatively simple, containing only one aspect word. Although there is noise interference from the word "best", the distance between the aspect word and the opinion word is relatively close, and all four groups of models made correct predictions.

In Sentence 2, there are two aspect words, "food" and "service", with opposite sentiment polarities. The aspect word "food" is close to the opinion word "Great" with less interference, so it is easy to determine the sentiment orientation of this aspect word. However, the aspect word "service" is easily misjudged as having the same sentiment polarity as "Great". Whereas baseline models like DualGCN primarily rely on dependency links, which may be misled by the contrastive conjunction "but", our model successfully captures the correct sentiment for both aspects. This can be attributed to the MLFF module, where the orthogonal projection of constituent features against the context helped isolate the contrastive phrase structure, and the cross-attention between dependency and knowledge features reinforced the strong negative sentiment clue "dreadful" specifically for the aspect "service".

In Sentence 3, there are only two words, but the emotional features of the keyword "Skip" were not captured by the other three models. On the contrary, our model introduced external knowledge and accurately predicted the sentiment polarity of the aspect word. In Sentence 4, the aspect word "Bombay" has multiple semantic information. With the help of background knowledge, our model handled complex and informal sentences well and made a correct prediction. The last example shows that, compared with positive and negative emotions, it is more difficult for many current models to identify the sentiment polarity of neutral sentences, which can be clearly seen in the specific experimental details. The model in this paper fully considered the complementarity of syntactic knowledge and semantic dependency information and used orthogonal projection and interactive attention mechanisms to process neutral sentences, achieving good classification accuracy.

4.9. Model visual analysis

In order to more accurately verify the performance of the model in specific cases, we conduct attention visualization analysis on six examples selected from the dataset, as shown in Figure 8.

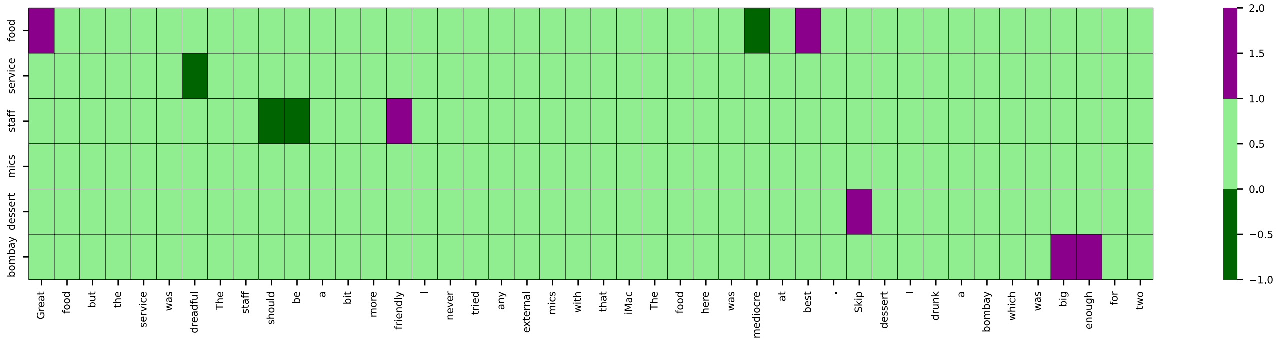

Figure 8. Model attention weight distribution.

During the experiment, as shown in Figure 8, our model can accurately capture the sentiment polarity of opinion words and assign corresponding sentiment score weights. When there are multiple opinion words with opposite emotions within the same lexical domain, such as in the sentence "The staff should be a bit more friendly", the words "should" and "be" are closer to negative emotions, while the adjective "friendly" is the opposite. Positive factors are presented, and through the intervention analysis of the model in this paper, the relatively negative emotional tendency of the aspect word "staff" can be well obtained. In specific ABSA tasks, sentiment analysis for neutral aspect words is the most difficult aspect. For example, in the sentence "I never tried any external mics with that iMac", it is necessary to determine the sentiment tendency of the aspect word "mics". However, the words in front of the sentences, such as "never" and "any external", all have some relatively negative characteristics. But the model in this paper starts from the sentence structure and dependency relationship, and through the fusion processing of the MLFF module, finally accurately predicts the sentiment of the aspect words.

5. Conclusions

In this paper, we present an ABSA model that integrates multi-granularity features and hierarchical networks. By integrating the constituent structure, semantic dependency information, external knowledge, and contextual feature information of the text, the model can obtain more comprehensive multi-granularity feature information. Additionally, we have designed an MLFF. By employing orthogonal projection technology and cross attention, and through iterative calculations, the MLFF can effectively harness the comprehensive potential and synergistic effects among features of different granularities. The results of extensive experiments demonstrate that our model can enhance the extraction and fusion of aspect-level features. Meanwhile, our model achieves superior experimental results when compared with other models based on LSTM, BERT, and RoBERTa. Future research will concentrate on investigating the noise and data redundancy issues arising from the introduction of external knowledge and actively exploring more effective multi-granularity feature extraction methods and fusion mechanisms.

Use of AI tools declaration

We declare that we have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Thanks to the National Natural Science Foundation of China (72174079) for its sponsorship and support of this research.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Z. Xu, Y. Wu, L. Tang, S. Gui, From user-generated content to quality improvement: A multi-granularity analysis of customer satisfaction and attention in new energy vehicles using deep learning, *Comput. Ind.*, **173** (2025), 104380. https://doi.org/10.1016/j.compind.2025.104380
2. X. Zhao, Y. Wang, Z. Zhong, Text-in-image enhanced self-supervised alignment model for aspect-based multimodal sentiment analysis on social media, *Sensors*, **25** (2025), 2553. https://doi.org/10.3390/s25082553
3. D. Tang, B. Qin, X. Feng, T. Liu, Effective LSTMs for target-dependent sentiment classification, in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, (2016), 3298–3307. https://doi.org/10.18653/v1/C16-1311
4. J. Wang, J. Li, S. Li, Y. Kang, M. Zhang, L. Si, et al., Aspect sentiment classification with both word-level and clause-level attention networks, in *Twenty-Seventh International Joint Conference on Artificial Intelligence*, (2018), 4439–4445. https://doi.org/10.24963/ijcai.2018 /617
5. R. Li, H. Chen, F. Feng, Z. Ma, X. Wang, E. Hovy, Dual graph convolutional networks for aspect-based sentiment analysis, in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, (2021), 6319–6329. https:// doi.org/ 10.18653/ v1/2021.acl-long.494
6. S. Liang, W. Wei, X. Mao, F. Wang, Z. He, BiSyn-GAT+: Bi-syntax aware graph attention network for aspect-based sentiment analysis, in *Findings of the Association for Computational Linguistics: ACL 2022*, (2022), 1835–1848. https://doi.org/10.18653/v1/2022.findings-acl.144
7. C. Zhang, Q. Li, D. Song, Aspectbased sentiment classification with aspect-specific graph convolutional networks, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, (2019), 4568–4578. https://doi.org/10.18653/v1/D19-1464
8. K. Sun, R. Zhang, S. Mensah, Y. Mao, X. Liu, Aspect-level sentiment analysis via convolution over dependency tree, in *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, (2019), 5679–5688. https://doi.org/10.18653/v1/D19-1569

9. B. Liang, R. Yin, L. Gui, J. Du, R. Xu, Jointly learning aspect-focused and inter-aspect relations with graph convolutional networks for aspect sentiment analysis, in *Proceedings of the 28th international Conference on Computational Linguistics*, (2020), 150–161. https://doi.org/10.18653 /v1/2020.coling-main.13
10. P. Zhao, L. Hou, O. Wu, Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification, *Knowl. Based Syst.*, **193** (2020), 105443. https://doi.org/10.1016/j.knosys.2019.105443
11. Y. Tian, G. Chen, Y. Song, Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble, in *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (2021), 2910–2922. https://doi.org/10.18653/v1/2021.naacl-main.231
12. J. Devlin, M. Chang, K. Lee, K. Toutanova, 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*, (2019), 4171–4186. https://doi.org/10.18653/v1/N19-1423
13. Z. Liu, W. Lin, Y. Shi, J. Zhao, A robustly optimized BERT pre-training approach with post-training, in *Proceedings of the 20th Chinese National Conference on Computational Linguistics*, (2021), 1218–1227.
14. Y. Wang, N. Yang, D. Miao, Q. Chen, Dual-channel and multi-granularity gated graph attention network for aspect-based sentiment analysis, *Appl. Intell.*, **53** (2023), 13145–13157. https://doi.org/10.1007/s10489-022-04198-5
15. S. Pang, Y. Xue, Z. Yan, W. Huang, J. Feng, Dynamic and multi-channel graph convolutional networks for aspect-based sentiment analysis, in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, (2021), 2627–2636. https://doi.org/10.18653/v1/2021.findings-acl.232
16. J. Zhou, J. X. Huang, Q. V. Hu, L. He, Sk-gcn: Modeling syntax and knowledge via graph convolutional network for aspect-level sentiment classification, *Knowl. Based Syst.*, **205** (2020),106292. https://doi.org/10.1016/j.knosys.2020.106292
17. Q. Zhong, L. Ding, J. Liu, B. Du, D. Tao, Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis, *IEEE Trans. Knowl. Data Eng.*, **35** (2023), 10098–10111. https://doi.org/10.1109/TKDE.2023.3250499
18. H. Liu, X. Li, W. Lu, K. Cheng, X. Liu, Graph augmentation networks based on dynamic sentiment knowledge and static external knowledge graphs for aspect-based sentiment analysis, *Expert Syst. Appl.*, **251** (2024), 123981. https://doi.org/10.1016/j.eswa.2024.123981
19. P. Mishra, S. K. Panda, Dependency structure-based rules using root node technique for explicit aspect extraction from online reviews, *IEEE Access*, **11** (2023), 65117–65137. https://doi.org/10.1109/ACCESS.2023.3287119
20. X. Shi, M. Hu, F. Ren, P. Shi, Prompted representation joint contrastive learning for aspect-based sentiment analysis, *Knowl. Based Syst.*, **285** (2024), 111345. https://doi.org/10.1016/j.knosys.2023.111345
21. M. Pota, M. Ventura, H. Fujita, M. Esposito, Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets, *Expert Syst. Appl.*, **181** (2023), 115119. https://doi.org/10.1016/j.eswa.2021.115119

22. C. Sun, L. Huang, X. Qiu, Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (2019), 380–385. https://doi.org/10.18653/v1/N19-1035
23. H. Xu, B. Liu, L. Shu, P. Yu, BERT post-training for review reading comprehension and aspect-based sentiment analysis, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (2019), 2324–2335. https://doi.org/10.18653/v1/N19-1242
24. L. Liu, Y. Yang, J. Hu, Simple but effective: A model for aspect-based sentiment analysis, in *Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020)*, (2020), 52–59. https://doi.org/10.1142/11911
25. K. Sun, R. Zhang, S. Mensah, Y. Mao, X. Liu, Aspect-level sentiment analysis via convolution over dependency tree, in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing* (EMNLP-IJCNLP), (2019), 5679–5688. https://doi.org/10.18653/v1/D19-1569
26. Z. Zhang, Z. Zhou, Y. Wang, SSEGCN: Syntactic and semantic enhanced graph convolutional network for aspect-based sentiment analysis, in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (2022), 4916–4925. https://doi.org/10.18653/v1/2022.naacl-main.362
27. X. Zhao, H. Peng, Q. Dai, X. Bai, H. Peng, Y. Liu, et al., RDGCN: Reinforced dependency graph convolutional network for aspect-based sentiment analysis, in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, (2024), 976–984. https://doi.org/10.1145/3616855. 3635775
28. D. Marcheggiani, I. Titov, Graph convolutions over constituent trees for syntax-aware semantic role labeling, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (EMNLP), (2020), 3915–3928. https://doi.org/10.18653/v1/2020. emnlp-main.322
29. Y. Li, C. Yin, S. Zhong, Sentence constituent-aware aspect-category sentiment analysis with graph attention networks, in *Natural Language Processing and Chinese Computing*, (2020), 815–827. https://doi.org/10.1007/978-3-030-60450-9_64
30. H. Fei, S. Wu, Y. Ren, F. Li, D. Ji, Better combine them together! integrating syntactic constituency and dependency representations for semantic role labeling, in *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, (2021), 549–559. https://doi.org/10.18653/v1/2021.findings-acl.49
31. Q. Zhong, L. Ding, J. Liu, B. Du, H. Xin, H. Tao, Knowledge graph augmented network towards multiview representation learning for aspect-based sentiment analysis, *IEEE Trans. Knowl. Data Eng.*, **35** (2023), 10098–10111. https://doi.org/10.1109/TKDE.2023.3250499
32. X. Shi, M. Hu, F. Ren, P. Shi, Prompted representation joint contrastive learning for aspect-based sentiment analysis, *Knowl. Based Syst.*, **285** (2024), 111345. https://doi.org/10.1016/j.knosys.2023.111345
33. J. Shi, W. Li, Q. Bai, Y. Yang, J. Jiang, Syntax-enhanced aspect-based sentiment analysis with multi-layer attention, *Neurocomputing*, **557** (2023), 126730. https://doi.org/10.1016/j.neucom. 2023.126730

34. X. Wang, M. Tang, T. Yang, Z. Wang, A novel network with multiple attention mechanisms for aspect-level sentiment analysis, *Knowl. Based Syst.*, **227** (2021), 107196. https://doi.org/10.1016/j.knosys.2021.107196
35. Q. Qin, W. Hu, B. Liu, Feature projection for improved text classification, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), 8161–8171. https://doi.org/10.18653/v1/2020.acl-main.726
36. X. Xue, C. Zhang, Z. Niu, X. Wu, Multi-level attention map network for multimodal sentiment analysis, *IEEE Trans. Knowl. Data Eng.*, **35** (2022), 5105–5118. https://doi.org/10.1109/TKDE.2022.3155290
37. M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, SemEval-2014 task 4: Aspectbased sentiment analysis, in *Proceedings of the 8thInternational Workshop on Semantic Evaluation (SemEval 2014)*, 27–35. https://doi.org/10.3115/v1/s14-2004
38. L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent twitter sentiment classification, in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, (2014), 49–54. https://doi.org/10.3115/v1/P14-2009
39. K. Wang, W. Shen, Y. Yang, X. Quan, R. Wang, Relational graph attention network for aspect-based sentiment analysis, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), 3229–3238. https://doi.org/ 10.18653/ v1/2020.acl-main.295
40. H. Tang, D. Ji, C. Li, Q. Zhou, Dependency graph enhanced dual-transformer structure for aspect-based sentiment classification, in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, (2020), 6578–6588. https://doi.org/10.18653/v1/2020.acl-main.588
41. G. Lu, J. Li, J. Wei, Aspect sentiment analysis with heterogeneous graph neural networks, *Inf. Process. Manage.*, **59** (2022), 102953. https://doi.org/10.1016/j.ipm. 2022.102953
42. Z. Xiao, J. Wu, Q. Chen, C. Deng, 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*, (2021), 9193–9200. https://doi.org/10.18653/v1/2021.emnlp-main.724
43. S. Feng, B. Wang, Z. Yang, J. Ouyang, Aspect-based sentiment analysis with attention-assisted graph and variational sentence representation, *Knowl. Based Syst.*, **258** (2022), 109975. https://doi.org/10.1016/j.knosys.2022.109975
44. R. Zhang, Q. Chen, Y. Zheng, S. Mensah, Y. Mao, Aspect-level sentiment analysis via a syntax-based neural network, *IEEE/ACM Trans. Audio Speech Lang. Process.*, **30** (2022), 2568–2583. https://doi.org/10.1109/TASLP.2022.3190731
45. Y. Wan, Y. Chen, L. Shi, L. Liu, A knowledge-enhanced interactive graph convolutional network for aspect-based sentiment analysis, *J. Intell. Inf. Syst.*, **61** (2023), 343–365. https://doi.org/10.1007/s10844-022-00761-1
46. T. Gu, H. Zhao, M. Li, Effective inter-aspect words modeling for aspect-based sentiment analysis, *Appl. Intell.*, **53** (2023), 4366–4379. https://doi.org/10.1007/s10489-022-03766-9
47. X. Xin, A. Wumaier, Z. Kadeer, J. He, SSEMGAT: Syntactic and semantic enhanced multi-layer graph attention network for aspect-level sentiment analysis, *Appl. Sci.*, **13** (2023), 5085. https://doi.org/10.3390/app13085085

48. J. Dai, H. Yan, T. Sun, P. Liu, X. Qiu, 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*, (2021), 1816–1829. https://doi.org/10.18653/v1/2021.naacl-main.145
49. Y. Cao, Y. Tang, H. Du, F. Xu, Z. Wei, C. Xin, Heterogeneous reinforcement learning network for aspect-based sentiment classification with external knowledge, *IEEE Trans. Affective Comput.*, **14** (2023), 3362–3375. https://doi.org/10.1109/TAFFC.2022.3233020
50. X. Shi, M. Hu, F. Ren, P. Shi, J. Deng, Y. Tang, Prompted and integrated textual information enhancing aspect-based sentiment analysis, *J. Intell. Inf. Syst.*, **62** (2024), 91–115.
51. X. Shi, M. Hu, F. Ren, P. Shi, Prompted representation joint contrastive learning for aspect-based sentiment analysis, *Knowl. Based Syst.*, **285** (2024), 111345. https://doi.org/10.1016/j.knosys.2023.111345