Expert Systems With Applications 255 (2024) 124523

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Label-semantics enhanced multi-layer heterogeneous graph convolutional network for Aspect Sentiment Quadruplet Extraction

Yiheng Fu [a], Xiaoliang Chen [a,b,c,*], Duoqian Miao [b], Xiaolin Qin [d], Peng Lu [c], Xianyong Li [a]

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

ARTICLE INFO

Dataset link: https://github.com/Thirring/LSEMH-GCN

Keywords:
Natural language processing
Aspect Sentiment Quadruplet Extraction
Graph convolutional network
Linguistic features
Label semantics

ABSTRACT

Aspect-based sentiment analysis (ABSA) is a pivotal area within natural language processing (NLP) that focuses on extracting fine-grained sentiment information from text. A particularly demanding task within ABSA is Aspect Sentiment Quadruplet Extraction (ASQE), which aims to identify quadruplets comprising aspect terms, their corresponding opinion terms, sentiment polarity, and categories. This level of analysis is crucial for downstream applications such as sentiment monitoring and user experience research, offering nuanced insights into textual data. However, current methodologies fall short of fully leveraging the rich linguistic features of sentences and the semantic information of labels embedded within sentences. To address this gap, this paper introduces a novel framework, the Label-Semantics Enhanced Multi-layer Heterogeneous Graph Convolutional Network (LSEMH-GCN), specifically designed for ASQE. Our model integrates sentence linguistic features and label semantics to construct a graph neural network tailored for this task. It employs a multi-layer graph convolutional network that synergizes various linguistic features, and utilizes Biaffine attention to enrich the label probability distribution for token pairs with semantic label information. Furthermore, our approach introduces a token pair vector concatenation strategy alongside an advanced asymmetric label tagging scheme to enhance quadruplet extraction. Comprehensive evaluations on benchmark datasets reveal that LSEMH-GCN significantly surpasses existing state-of-the-art models, establishing a new benchmark for ASQE. Our model achieves an average F1 score improvement of 15.52% and 12.30% on the Restaurant-ACOS and Laptop-ACOS datasets, respectively.

1. Introduction

Aspect-Based Sentiment Analysis (ABSA) refines sentiment analysis by pinpointing and evaluating sentiments tied to specific aspect terms in texts. This extension enriches sentiment analysis by revealing detailed perceptions that surpass the general sentiment classification at the sentence level, thus enabling a more profound comprehension of textual sentiments. ABSA's advanced analytical prowess is pivotal for various applications, including recommendation systems (Chen et al., 2023), question-answering systems (Qiu et al., 2021), and sentiment-driven dialogue systems (Wei et al., 2021). Traditionally, ABSA research (Lu et al., 2022; Yu et al., 2022; Zhou et al., 2023a) has concentrated on identifying aspect terms and their sentiment polarities, providing a foundational yet somewhat restricted view of sentiment's complexity in texts. Recent advancements (Cai et al., 2021; Xiong et al., 2023; Zhou et al., 2023b) in the field have evolved to address this limitation by aiming to identify four key aspect-level sentiment components: aspect terms, opinion terms, aspect categories, and sentiment polarities. Termed Aspect Sentiment Quadruplet Extraction (ASQE), this new challenge presents a significant departure from traditional methods, offering a more comprehensive and nuanced exploration of textual sentiments.

Fig. 1 illustrates the structured methodology of ASQE, providing a concrete example of how sentiments are dissected and interpreted in a detailed manner. It showcases a sentence and its dependency tree structure, highlighting the intricate relationship between various linguistic elements. Aspect terms and opinion terms are highlighted in blue and yellow, respectively, underscoring their crucial roles in forming sentiment meanings. Aspect categories, shown in purple, represent thematic classifications within the text, while sentiment polarities, marked in green, indicate the sentiment's valence associated with each aspect.

* Corresponding author at: School of Computer and Software Engineering, Xihua University, Chengdu, 610039, PR China.
E-mail address: chenxl@mail.xhu.edu.cn (X. Chen).

https://doi.org/10.1016/j.eswa.2024.124523
Received 3 April 2024; Received in revised form 16 May 2024; Accepted 15 June 2024
Available online 19 June 2024

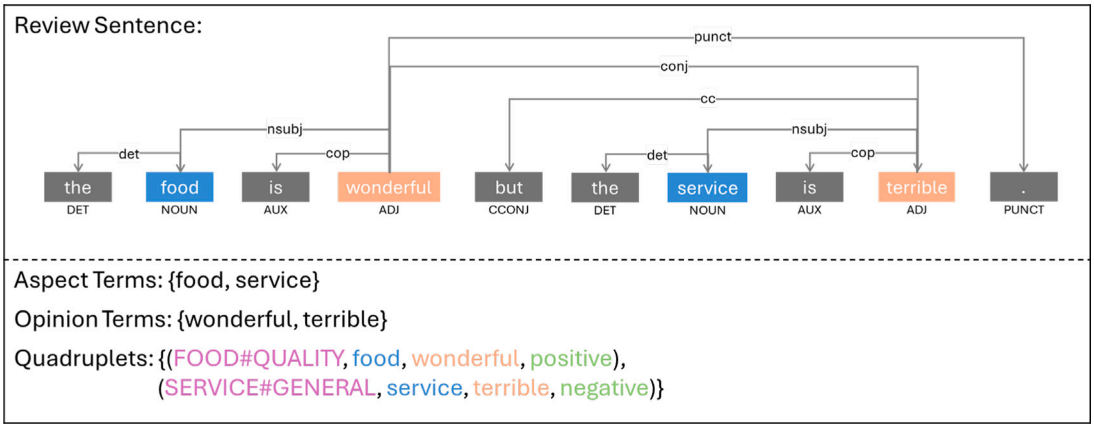

Fig. 1. Dependency tree visualization of ASQE with color-coded sentiment elements.

Historical ABSA research segmented the analysis into multiple subtasks, focusing on predicting individual sentiment elements, such as Aspect Target Extraction (ATE) (Ma et al., 2019), Opinion Target Extraction (OTE) (Fan et al., 2019), Aspect Category Detection (ACD) (Hu et al., 2021), and Aspect Sentiment Classification (ASC) (Ma et al., 2017). However, recent studies have shifted towards the simultaneous extraction of multiple sentiment elements. Notably, the Aspect-Opinion Pair Extraction (AOPE) (Chen et al., 2020a) task identifies aspect and opinion terms together, while Aspect Sentiment Triplet Extraction (ASTE) (Xu et al., 2021) aims to isolate triplets of aspect, opinion, and sentiment, showcasing the field's growing complexity and analytical capabilities.

Despite the strides made in ABSA, existing research primarily focuses on the isolated extraction of sentiment elements, falling short of capturing the comprehensive aspect-level sentiment architecture. To bridge this gap, the ASQE task emerges, aiming to simultaneously predict four sentiment elements within a quadruplet structure. This approach, as illustrated in Fig. 1, analyzes the sentence "*the food is wonderful but the service is terrible.*" to generate distinct quadruplets, such as (FOOD#QUALITY, food, wonderful, positive) and (SERVICE#GENERAL, service, terrible, negative). This method not only enhances the granularity of sentiment analysis but also significantly boosts the accuracy of linking sentiments to their respective aspects, thereby propelling the field forward.

Pioneering work by Cai et al. (2021) laid the groundwork for the ASQE task, establishing baseline models that have become benchmarks in the field. Subsequent research by Zhang et al. (2021a) and Bao et al. (2022) delved into generative methodologies for ASQE, with further expansions by Gou et al. (2023). Ye et al. (2023) applied machine reading comprehension strategies to tackle the ASQE challenge innovatively. Moreover, studies (Chen et al., 2022; Li et al., 2021) have demonstrated that models incorporating linguistic features to construct graph convolutional networks (GCN) show promising results, highlighting the importance of linguistic attributes in boosting ABSA performance. However, the application of linguistic features to the ASQE task remains an uncharted territory, presenting challenges that our study seeks to address:

- **How to unlock the potential of linguistic features for enhanced ASQE model performance ?**
  Integrating linguistic features into the ASQE task poses a significant challenge. For instance, aspect terms shown in Fig. 1 like "food" and "service" are nouns, while opinion terms like "wonderful" and "terrible" are adjectives, all connected by the dependency relationship "nsubj" within the dependency tree structure. These linguistic features are crucial for extracting and associating aspect and opinion terms (Chen et al., 2022). However, their application in ASQE is complicated by implicit representations, where crucial elements within aspect-opinion pairs are not explicitly mentioned. The limitations of current syntactic parsers, which map syntactic connections among explicit words and assign part-of-speech tags, hinder the direct application of linguistic features to ASQE. Addressing this challenge requires a method that uncovers hidden elements and establishes structured connections between them and the explicit words in sentences.
- **How to elevate ASQE by advanced multi-class classification techniques ?**
  ASQE evolves from the conventional ASTE by emphasizing the prediction of aspect categories, a critical factor in the model's precision and overall performance. Despite the advantages of using linguistic features for identifying aspect-opinion pairs, a gap remains in correlating these features with the complex task of category classification. The vast array of categories, exemplified by the commonly used Laptop-ACOS dataset's 121 categories (Cai et al., 2021), underscores the classification process's complexity and the requirement for an innovative classification optimization strategy.

This work introduces the Label-Semantics Enhanced Multi-layer Heterogeneous Graph Convolutional Network (LSEMH-GCN), a novel methodology designed to navigate the complexities previously identified. Our approach is segmented into three pivotal components, each playing a crucial role in enhancing the model's novelty and efficacy.

First, our research introduces a Linguistic Feature Module (LFM) designed to address the ASQE challenge. This module lies on the strategic utilization of sentence-level feature representations, a method crafted to unveil the subtle implicit meanings and intricate structures through linguistic features. Central to this module is the integration of four key linguistic features: token pair part-of-speech combinations, syntactic dependency types, distances derived from constituent tree structures, and relative positional distances. This multifaceted framework is meticulously engineered to enable our model to capture the nuanced implicit representations and meaningful structures that exist among a diverse vocabulary. These elements are adeptly transformed into graph edges, with token hidden representations forming the graph nodes. Then, utilizing Graph Convolutional Networks (GCN), we achieve a novel standard in aggregating information for linguistic feature integration within graph neural networks.

Second, a designed Label Semantic Module (LSM) introduces an innovative application of Biaffine attention, adept at capturing the semantic subtleties of categories and sentiment labels. This module excels in classifying token pairs and distributing sentiment labels, generating a label probability tensor for each token pair with unparalleled detail and specificity.

Third, a Quadruplet Extraction Module (QEM) is introduced. This module represents a methodological breakthrough by combining token pair representations with different token pair relationships using a novel technique of vector concatenation. This technique skillfully separates the vectors representing category classifications from those representing sentiment relations, culminating in the adoption of an asymmetric quadruplet labeling scheme for quadruplet extraction. Our research introduces methodological contributions in ASQE solutions, as outlined below:

- **Pioneering Graph Convolutional Networks Integration:** We introduce the first application of GCN to ASQE, establishing a groundbreaking methodology that expands the potential for future ASQE research. This approach sets a new benchmark for the domain.
- **Semantic Features for Enhanced Classification:** Our novel category classification technique, which capitalizes on the semantic features of category labels, significantly refines the precision of both category and sentiment classification within ASQE tasks. This achievement demonstrates our capability to advance classification methodologies through semantic analysis.
- **Concrete Implicit Feature Representation:** By incorporating various label and positional information settings, we address the challenge of implicit representation in ASQE. Our approach strengthens these representations, linking them with existing words and effectively integrating multiple linguistic features into ASQE's framework. This advancement not only resolves a pivotal challenge but also broadens the applicability of ASQE models.
- **Quadruplet Extraction with Vector Concatenation:** Our unique vector concatenation technique, combined with an asymmetric quadruplet labeling scheme, enables efficient and accurate end-to-end quadruplet extraction. This development signifies our dedication to advancing the capabilities of ASQE.

2. Related work

Our exploration delves into the intricate realm of Aspect-Based Sentiment Analysis (ABSA), a sector increasingly influenced by the advent of deep learning techniques. The application of these methodologies within ABSA can be broadly categorized into two distinct streams: those that leverage syntactic information (Syntax-based methods) and those that do not (Syntax-free methods).

Syntax-based methods: The incorporation of linguistic features into the architecture of models has been at the forefront of several groundbreaking studies within ABSA. Notably, Li et al. (2021) innovatively introduced a dual graph convolutional network that harmonizes syntactic dependency trees with semantic dependency trees, presenting a sophisticated approach to analyzing sentiment polarity. Similarly, Liang et al. (2022) pioneered the exploration of constituent trees to address the challenge posed by the proximity of opinion terms to unrelated aspects in syntactic dependency trees, marking a significant and beneficial advancement in the field. Their innovation significantly enhances the model's ability to discern relevant linguistic relationships, as demonstrated in Fig. 2, by effectively increasing the relative distance between vocabularies associated with different aspects and reducing potential interference. Building on these advancements, Chen et al. (2022) ingeniously integrated linguistic features into a multi-channel graph convolutional network, applying this novel structure to the task of triplet extraction. These studies highlight the dynamic progression of ABSA and underscore the profound impact of syntax-based methods in enhancing the precision and depth of sentiment analysis.

Syntax-free methods: In the realm of ABSA, various methodologies that operate independently of linguistic features have made significant strides, particularly in recent ABSA research focusing on the ASQE task. ASQE stands out as the most complex and demanding task within ABSA, primarily due to its requirement to capture and integrate a broad spectrum of sentiment elements comprehensively. In a seminal contribution to this field, Cai et al. (2021) introduced two new datasets annotated with sentiment quadruplets. They also established a series of pipeline baselines by integrating existing models, thus laying the groundwork for future research in ASQE. Expanding upon this groundwork, Zhang et al. (2021a) advanced the field by introducing a PARAPHRASE modeling paradigm aimed at transforming the prediction of sentiment quadruplets into a paraphrase generation process. This end-to-end approach reduces error propagation in pipeline methods and allows models to consider the semantic information of sentiment elements through natural language generation. Furthermore, Bao et al. (2022) introduced a novel opinion tree generation model that highlights the semantic relationships of sentiment elements (e.g., aspect terms, opinion terms), revealing a more comprehensive and complete aspect-level semantic structure for extracting sentiment elements. They proposed two strategies, constrained decoding algorithms and sequence-to-sequence joint learning for pre-trained tasks, to effectively form opinion tree structures. Additionally, Ye et al. (2023) addressed the strong interdependence of subtasks within the ASQE task by transforming ASQE into a multi-turn MRC task, enabling models to learn relationships between subtasks effectively. They also proposed a hierarchical category classification strategy to handle complex category scenarios, utilizing bidirectional attention mechanisms to enhance context representation.

While existing research in ASQE has made significant progress, the incorporation of linguistic features remains untapped. The utilization of linguistic features can provide a more systematic approach to investigating ABSA, supported by a wealth of studies showcasing their efficacy, motivating us to incorporate linguistic features into the ASQE task. The application of linguistic features can introduce syntax analysis techniques and a more interpretable perspective to ASQE research, expanding the methods and prospects of this field. Therefore, our study charts a new course. We harness the power of graph convolutional networks to dissect and analyze linguistic features, creating meaningful structured relationships between tokens. This approach not only enhances our understanding of linguistic structures but also seamlessly integrates these insights into the ASQE task.

3. Methodology

In this section, we introduce a novel model that harnesses the capabilities of GCNs combined with deep linguistic insights to perform ASQE. Our proposed model, named the Label-Semantics Enhanced Multi-layer Heterogeneous Graph Convolutional Network (LSEMH-GCN), is meticulously crafted to comprehensively extract structured sentiment quadruplets (category, aspect, opinion, sentiment polarity) from the input text, enabling a rich representation of the multifaceted sentiment dynamics.

The LSEMH-GCN model distinguishes itself by its innovative integration of enhanced label semantics and linguistic features within a framework of multi-layered heterogeneous graph convolutions. This integration is pivotal in modeling the complex interplay between words, aspects, opinions, categories, and sentiment polarities. By doing so, our approach not only captures the essence of sentiment relationships but also maps these relationships within a structured analytical framework. The framework, as depicted in Fig. 3, is poised to set new benchmarks in performance, extending its utility across a broad range of sentiment-driven applications and domains.

3.1. Problem definition

In the realm of ABSA, with a focus on ASQE, we consider a sentence $W = \{w_1, w_2, \dots, w_n\}$ composed of n words. The challenge of ASQE lies in the precise identification and extraction of a set of sentiment quadruplets, denoted as $Q = \{(c, a, o, s)_m\}_{m=1}^{|Q|}$. Each quadruplet (c, a, o, s) is a structured entity that captures four essential dimensions of sentiment analysis:

- a signifies the aspect term under scrutiny, referring to the subject or feature within the sentence that is the target of the sentiment.

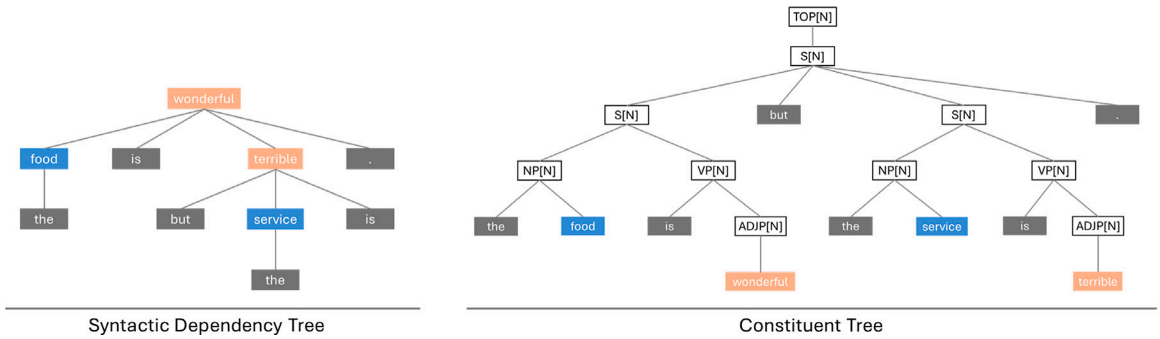

Fig. 2. Depicting the dependency and constituent trees for a same example sentence.

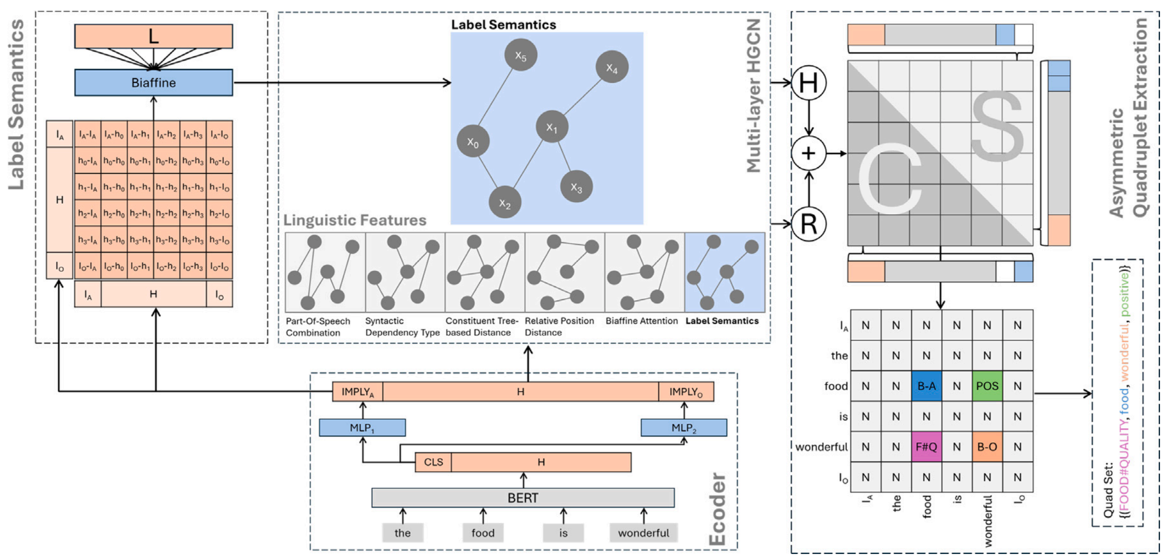

Fig. 3. Overall architecture of our end-to-end model LSEMH-GCN.

- o denotes the opinion term associated with the aspect term a, representing the specific sentiment expression or evaluation about the aspect.
- c identifies the category of the aspect-opinion pair (a, o), a classification that varies based on the dataset, illustrating the diverse nature of sentiment analysis across different sectors.
- s indicates the sentiment polarity linked to the aspect-opinion pair (a, o), categorized into one of three types: positive (POS), for favorable sentiments; neutral (NEU), indicating the absence of clear sentiment; or negative (NEG), for adverse sentiments.

The quantity of such quadruplets within a sentence is represented by $|Q|$. This detailed approach to sentiment analysis via ASQE aims to enrich our comprehension of the intricate relationship between aspects and opinions in text, thereby enhancing the accuracy of sentiment analysis by capturing the complexity of human emotions and evaluations as conveyed through language.

3.2. Label-Semantics Enhanced Multi-layer Heterogeneous Graph Convolutional Network model

3.2.1. Overview

This section provides a comprehensive overview of the architectural framework underpinning the proposed Label-Semantics Enhanced Multi-layer Heterogeneous Graph Convolutional Network (LSEMH-GCN) model, specifically designed for tackling ASQE tasks. The LSEMH-GCN architecture is meticulously constructed around five pivotal components, each contributing uniquely to the model's efficacy in ASQE:

- **Input and Encoding Layer:** This foundational layer is tasked with the initial processing of the input sentence, employing advanced encoding techniques to prepare the groundwork for subsequent ASQE analysis. Contextual embeddings are leveraged to ensure a rich representation of the input's semantic intricacies.
- **Enhanced Label Semantics Module:** At the core of our LSEMH-GCN model, this module improves the text sentiment analysis by adding sentiment polarity and category-specific descriptive labels. Utilizing Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) for encoding and the Biaffine Attention mechanism, it categorizes sentiment into positive, neutral, and negative, while also recognizing non-sentimental or categorical elements. This process turns labels into a set of features that capture the sentiment's nuances. The Biaffine Attention evaluates how token pairs relate to these labels, creating a detailed view of sentiment and category dynamics, significantly improving sentiment analysis accuracy.
- **Linguistic Feature Module:** This component integrates a diverse array of linguistic features, including syntactic, positional, and dependency-related attributes, into the LSEMH-GCN's analytical framework. The model attains a comprehensive understanding of sentence structure and the dynamic interplay among textual elements by embedding these linguistic insights.
- **Multi-layer Heterogeneous Graph Convolutional Network:** Drawing inspiration from Convolutional Neural Networks (CNNs) yet tailored for graph-based data, this component introduces a

multi-layer heterogeneous graph convolutional network. It focuses on efficient information aggregation and the nuanced modeling of sentence structure through direct edges, enabling the model to capture complex relationships within the data.
- **Asymmetric Quadruplet Extraction Strategy:** This strategy refines the model's approach to the identification and extraction of category and sentiment. By employing a quadruplet tagging mechanism, the model enhances its precision in distinguishing between sentiment and category through specific vector concatenation techniques.

3.2.2. Input and encoding layer

In this layer, we employ the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) as our foundational encoding mechanism. This choice allows us to convert the input sentence into a detailed set of contextualized feature representations. For an input sentence $W = \{w_1, w_2, \dots, w_n\}$, the encoding layer generates a sequence of hidden feature representations $H = \{[CLS], h_1, h_2, \dots, h_n\}$. Here, the token $[CLS]$ serves as a sentence-level feature aggregator, encapsulating the comprehensive information of the entire sentence, while each h_i corresponds to the feature representation of the individual word w_i. To address the challenge of capturing connections within the sentence and between words and their implicit meanings, especially when aspects and opinions are implied rather than explicitly stated, we adopt a specific approach. We use a linear transformation of the sentence-level $[CLS]$ token to generate initial implicit representations for aspects (I_A) and opinions (I_O). These representations are then positioned at the start and end of the sequence of word feature representations, leading to an enhanced sequence $X = \{x_0, x_1, x_2, \dots, x_n, x_{n+1}\}$ that includes the implicit representations. The augmentation process is defined as follows:

$$x_0 = I_A = \mathrm{MLP}_1([CLS]) \tag{1}$$

$$x_{n+1} = I_O = \mathrm{MLP}_2([CLS]) \tag{2}$$

$$x_i = h_i, i \in [1, n] \tag{3}$$

This setup lays the groundwork for the model to effectively perform sentiment quadruplet extraction.

3.2.3. Enhanced label semantics module

The enhanced label semantics module introduces a refined labeling strategy to accurately depict sentiment polarities and categories in the text. We categorize sentiments into three labels: $\{positive\ emotion, neutral\ emotion, negative\ emotion\}$ to cover the range of emotional responses. Additionally, we use a label $\{it\ is\ not\ emotion\ or\ category\}$ for tokens that do not express sentiment or belong to any category, alongside specific category labels such as $\{food\ quality\}$. This creates a detailed label set that captures both sentiment polarity and category aspects. Leveraging the BERT model, we transform these labels into a set of hidden feature representations $L = \{l_1, l_2, \dots, l_m\}$, where m is the number of unique labels, embedding complex label semantics. To enhance the semantic interpretation of these labels concerning the text, we employ a Biaffine attention mechanism (Dozat & Manning, 2017). This mechanism calculates the probability distribution of sentiment and category labels across pairs of tokens. The process is mathematically described as follows:

$$Y = \left\{ y_{ij} \mid y_{ij} = x_i \oplus x_j, x_i \in X, x_j \in X, i, j \in [0, n+1] \right\} \tag{4}$$

$$g_{ij,k} = y_{ij}^{\mathrm{T}} U_1 l_k + U_2 \left(y_{ij} \oplus l_k \right) + b \tag{5}$$

$$r_{ij,k} = \frac{\exp\left(g_{ij,k}\right)}{\sum_{p=1}^{m} \exp\left(g_{ij,p}\right)} \tag{6}$$

$$R = \mathrm{Biaffine}(Y, L) \tag{7}$$

Here, $r_{ij,k}$ represents the probability that the token pair (x_i, x_j) is associated with label l_k. The vector $r_{ij} \in \mathbb{R}^{1\times m}$ indicates the probability distribution for assigning sentiment polarity and category labels to the token pair (x_i, x_j), with m reflecting the total number of labels. $R \in \mathbb{R}^{(n+2)\times(n+2)\times m}$ encapsulates the label probability distribution for all token pairs, with its third dimension indexing the m label types. Parameters U_1, U_2, and b are the learnable weights and biases within the model, and $\oplus$ signifies the operation of concatenation. The comprehensive equation, encapsulated in Eq. (7), is derived from the integration of Eqs. (4) to (6), illustrating the methodological precision in calculating label probabilities for enhanced semantic analysis.

3.2.4. Linguistic features

Our methodology integrates a comprehensive suite of four linguistic features into the ASQE framework, enhancing its capability to capture the nuanced semantic relationships within text. These features include:

- Part-of-Speech (POS) Pairs: The model is enhanced with POS tagging, incorporating unique labels I_A and I_O to implicit represent aspect and opinion terms, respectively. This addition enables a more detailed comprehension of the grammatical functions fulfilled by implicit representations in expressing sentiment.
- Syntactic Dependency Relations: To further dissect the structural intricacies of sentences, we incorporate syntactic dependency relations, including the "of" relation, I_{A2O} (aspect to opinion) relation, and I_{O2A} (opinion to aspect) relation. These dependencies are crucial for identifying the interconnections between aspect and opinion terms, thereby enriching the model's interpretative depth.
- Relative Distances Based on Constituent Trees: This feature quantifies the relative distances between elements in a sentence as determined by constituent tree structures. By setting the distance between implicit representations (e.g., I_A and I_O) and other words to 1, we provide a simplified yet effective metric for gauging proximity within the sentence structure, as depicted in Fig. 4.
- Relative Distances Within Sentences: Complementing the above, this feature captures the linear distances between words within sentences, offering an additional layer of contextual information that aids in the accurate mapping of semantic relationships.

By utilizing these linguistic features, a learnable embedding table is utilized to convert them into adjacency tensors. These tensors form the basis for graph convolution operations, allowing the model to seamlessly incorporate and analyze various levels of linguistic data.

3.2.5. Multi-layer heterogeneous graph convolutional network

This subsection explores the application of Graph Convolutional Networks (GCNs), an adaptation of Convolutional Neural Networks (CNNs) designed for graph-structured data. Unlike CNNs, which excel with grid-like data, GCNs are adept at handling data represented as graphs, comprising nodes and edges. They perform convolution operations across the graph, aggregating information from neighboring nodes to capture the data's structural essence. In sentences with n words, we typically construct an adjacency matrix $A \in \mathbb{R}^{n\times n}$ based on the syntactic dependency tree to represent the sentence's graph structure. In this matrix, the entry A_{ij} signifies the presence of an edge between the node pair (w_i, w_j). Here, an entry A_{ij} indicates an edge between nodes (w_i, w_j), with a value of 1 for direct connections and 0 otherwise. Recent innovations (Chen et al., 2020b; Guo et al., 2019; Li et al., 2021) have introduced soft edges, utilizing attention mechanisms to assign probabilistic weights to edges. This approach offers a more nuanced representation of the relevance between node pairs (w_i, w_j).

Our model integrates the enhanced label semantics module and the linguistic feature module to determine the probability distribution of relationships between token pairs. This is achieved using Biaffine attention, which models the relationships between all token pairs (Chen

Table 1
Multiple token pair tags are depicted in the figure, with LF denoting the Language Feature Module and LS representing the Enhanced Label Semantics module. Tags starting from the 13th item onwards correspond to distinct categories, the specifics of which are dictated by the dataset's content and quantity.

	Tag	LF	LS	Meaning
1	N	✓	✓	The token pair is not related to sentiment or categories.
2	B-A	✓		The token marks the beginning of an aspect term.
3	I-A	✓		The token marks the inside of an aspect term.
4	A	✓		Tokens (x_i, x_j) belongs to the same aspect term.
5	B-O	✓		The token marks the beginning of an opinion term.
6	I-O	✓		The token marks the inside of an opinion term.
7	O	✓		Tokens (x_i, x_j) belongs to the same opinion term.
8	S	✓		The aspect-opinion pair represented by the token pair conveys sentiment but does not specify the polarity.
9	POS		✓	The sentiment to be conveyed by the aspect-opinion pair that the token pair represented.
10	NEU		✓	
11	NEG		✓	
12	C	✓		The aspect-opinion pair represented by the token pair exists within a category, but does not differentiate specific categories.
⋮	⋮		✓	The category to which the aspect-opinion pair represented by the token pair belongs.(i.e., FOOD#QUALITY)

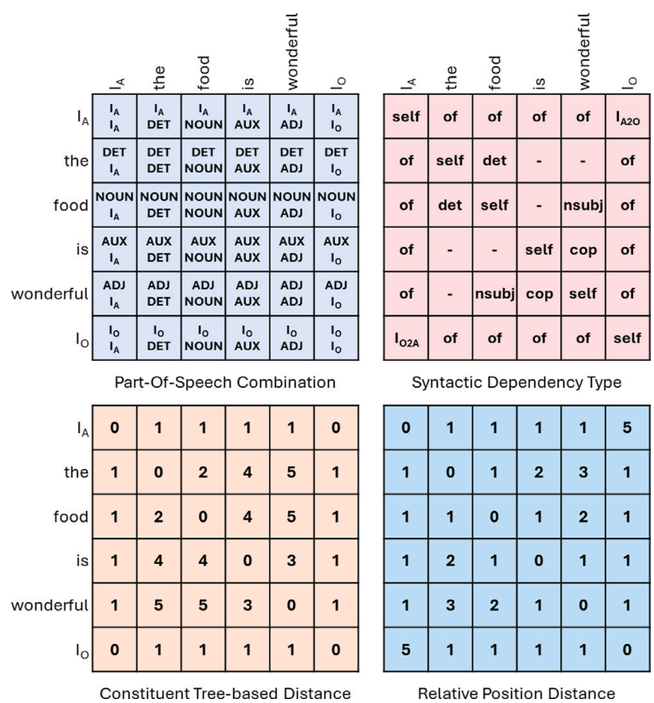

Fig. 4. Four distinct linguistic features.

et al., 2022), leading to the creation of multi-layered adjacency matrices. Each layer in these matrices represents a distinct relationship type between token pairs, as detailed in Table 1.

Subsequently, the model applies GCN to aggregate information across nodes, using the adjacency matrices to guide the convolution operations. For example, in the enhanced label semantics module, the aggregation process is described as follows:

$$\widetilde{X}_k^{ls} = \sigma\left(R^{ls}_{:,:,k} X W_k + b_k\right) \tag{8}$$

$$\widetilde{X}^{ls} = f(\widetilde{X}_1^{ls}, \widetilde{X}_2^{ls}, \ldots, \widetilde{X}_m^{ls}) \tag{9}$$

Here, $R^{ls}_{:,:,k}$ denotes the kth layer of the adjacency matrix R^{ls}, derived from the enhanced label semantics module. The parameters W_k and b_k represent learnable weights and biases, respectively. The function σ typically refers to the ReLU activation function, and $f(\cdot)$ is an average pooling operation that combines the token hidden representations from all convolutional layers, ensuring a comprehensive integration of information.

3.2.6. Asymmetric quadruplet extraction strategy

This section extends the methodology of triplet labeling extraction, as proposed by Chen et al. (2022), to introduce a novel quadruplet labeling extraction strategy tailored for the nuanced demands of ASQE. This strategy utilizes a set of tags $\{B\text{-}A, I\text{-}A, A\}$ for demarcating aspect terms and $\{B\text{-}O, I\text{-}O, O\}$ for opinion terms. Additionally, sentiment polarities of aspect-opinion pairs are encoded using $\{POS, NEU, NEG\}$ tags, while specific category tags are utilized for classifying these pairs. An illustrative example presented in Fig. 5 demonstrates how the $B\text{-}A$ tag and related tags identify an aspect term, $B\text{-}O$ and associated tags pinpoint opinion terms, and the POS tag signifies a positive sentiment polarity for the aspect-opinion pairs (wine list, interesting) and (wine list, good values). The tags D#S and D#P further categorize these pairs under DRINKS#STYLE_OPTIONS and DRINKS#PRICES, respectively, enabling the extraction of two distinct quadruplets: (DRINKS#STYLE_OPTIONS, wine list, interesting, positive) and (DRINKS#PRICES, wine list, good values, positive).

To accommodate the asymmetric nature of this labeling scheme, we propose a specialized vector concatenation technique that precisely differentiates between vectors representing the sentiment of aspect-opinion pairs and those denoting categories. For vectors in the lower triangular region of the token pair matrix, we append the category label probabilities from the label probability vector. Conversely, for vectors in the upper triangular region, we append the sentiment label probabilities. The token pair representation z_{ij} is then processed through a linear layer and a softmax function to yield a label probability distribution p_{ij}. The formulation is as follows:

$$\overline{\mathbf{X}} = f\left(\widetilde{X}^{\mathrm{psc}}, \widetilde{X}^{\mathrm{dep}}, \widetilde{X}^{\mathrm{ctd}}, \widetilde{X}^{\mathrm{rpd}}, \widetilde{X}^{\mathrm{ba}}, \widetilde{X}^{\mathrm{ls}}\right) \tag{10}$$

$$\mathbf{R} = R^{psc} \oplus R^{dep} \oplus R^{ctd} \oplus R^{rpd} \oplus R^{ba} \tag{11}$$

$$z_{ij} = \begin{cases} \overline{x}_i \oplus \overline{x}_j \oplus r_{ii} \oplus r_{jj} \oplus r_{ij} \oplus r^{ls}_{ij,1} \oplus [0]_3 \oplus r^{ls}_{ij,5:} & , i > j \\ \overline{x}_i \oplus \overline{x}_j \oplus r_{ii} \oplus r_{jj} \oplus r_{ij} \oplus r^{ls}_{ij} & , i = j \\ \overline{x}_i \oplus \overline{x}_j \oplus r_{ii} \oplus r_{jj} \oplus r_{ij} \oplus r^{ls}_{ij,:4} \oplus [0]_{len(r^{ls}_{ij})-4} & , i < j \end{cases} \tag{12}$$

$$p_{ij} = \mathrm{softmax}(W_p z_{ij} + b_p) \tag{13}$$

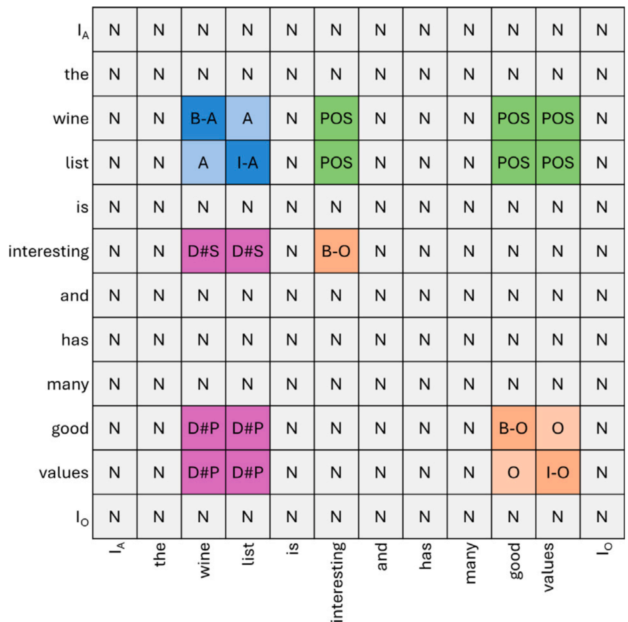

Fig. 5. Illustration of quadruplet tag assignment, where each cell denotes the tag assigned to a token pair. For detailed explanations of each tag, please consult Table 1.

In the above formulas, $f(\cdot)$ denotes an average pooling function. $\overline{X} = \{\overline{x}_0, \overline{x}_1, \ldots, \overline{x}_{n+1}\}$ and $\mathbf{R} = \{r_{00}, r_{01}, \ldots, r_{(n+1)(n+1)}\}$ represent the aggregated token feature representations and the probability distribution of token pair relationships, respectively. z_{ij} is the final representation of the token pair (x_i, x_j). The notation $r^{ls}_{ij,1}$ refers to the first dimension of r^{ls}_{ij}, $[0]_m$ is a zero vector of dimension m, and $r^{ls}_{ij,5:}$ and $r^{ls}_{ij,:4}$ denote specific segments of r^{ls}_{ij}. The dimensionality of r^{ls}_{ij} is indicated by $len(r^{ls}_{ij})$. W_p and b_p are learnable parameters, and p_{ij} represents the probability distribution across all labels, excluding the **S** and **C** labels as detailed in Table 1.

Algorithm 1 delineates the systematic approach for extracting quadruplets, which encapsulates the intricate process of identifying and correlating aspect and opinion terms with their respective sentiment and category labels within a sentence. The algorithm commences by extracting the principal diagonal from the label prediction matrix P, which facilitates the identification of aspect and opinion term spans. Subsequently, for each potential pairing of aspect and opinion terms, the algorithm scrutinizes the corresponding intersection area within P to catalog all encountered sentiment and category labels. The absence of either sentiment or category labels within this intersection signifies the infeasibility of forming a valid quadruplet. In scenarios where multiple labels of the same type (e.g., several sentiment or category labels) are present, the algorithm opts for the label exhibiting the highest occurrence frequency to constitute the quadruplet. This methodology is consistently applied to extract all viable quadruplets from the sentence.

In the context of analyzing the time complexity of the Quadruplet Extraction Strategy as outlined in the provided Algorithm 1, it is essential to dissect the algorithm into its constituent operations and assess the computational cost associated with each. The algorithm's primary objective is to extract quadruplets from a sentence based on the label prediction matrix P and the number of labels $label\_num$. The analysis below assumes that the sentence contains n tokens, leading to a $n \times n$ label prediction matrix P.

- Initialization: The algorithm begins by initializing an empty list $Quads$ and performing a series of retrieval operations to extract the main diagonal D, aspect spans A, and opinion spans O from P. The time complexity of these initial steps is largely dependent on the implementation of the GetDiagonal, GetAspectSpan, and GetOpinionSpan functions. Assuming efficient implementations, the extraction of the diagonal and identification of spans can be accomplished in $O(n)$ time, where n is the number of tokens in the sentence.
- Nested Loops: The core of the algorithm involves nested loops iterating over aspect spans A and opinion spans O. Let a and o denote the number of aspect and opinion spans, respectively. The worst-case scenario occurs when every token is identified as part of an aspect or opinion span, leading to $a, o = O(n)$.
- Label Counting and Quadruplet Formation: Within the nested loops, the algorithm iterates over the intersection areas defined by aspect and opinion spans in the matrix P, counting the occurrence of each label. Since the label counting involves accessing elements within a slice of the matrix and incrementing counters, the time complexity for this step is proportional to the size of the intersection area. In the worst case, where an aspect span or opinion span encompasses the entire sentence, this step could take $O(n^2)$ time due to the need to iterate over all elements in the slice. However, in practice, aspect and opinion spans are likely to be much smaller than the entire sentence, reducing the practical time complexity of this step.
- Label Selection and Quadruplet Appending: After counting label occurrences, the algorithm selects the labels with the highest frequency for sentiment and category, and appends the resulting quadruplet to the list $Quads$. The selection of maximum labels can be done in $O(label\_num)$ time, and appending to the list is $O(1)$.

Algorithm 1 Quadruplet Extraction Strategy

```
Require: The label prediction matrix P of a sentence and the number
    of labels label_num.
Ensure: A list of quadruplets Quads.
function QUADEXTRACTION(P, label_num)
    Quads ← []
    D ← GetDiagonal(P)
    A ← GetAspectSpan(D)
    O ← GetOpinionSpan(D)
    for as, ae in A do
        for os, oe in O do
            label_count ← [0] ∗ label_num
            for label in P[as : ae, os : oe] do
                label_count[label] + +
            end for
            for label in P[os : oe, as : ae] do
                label_count[label] + +
            end for
            if Sum(label_count[sentiment])= 0 or
    Sum(label_count[category])= 0 then
                continue
            end if
            s_label ← MaxLabel(label_count[sentiment])
            c_label ← MaxLabel(label_count[category])
            Quads.append([c_label, (as, ae), (os, oe), s_label])
        end for
    end for
    return Quads
end function
```

Given the above considerations, the overall time complexity of the algorithm can be expressed as $O(a \cdot o \cdot n^2 + a \cdot o \cdot label\_num)$. In the worst-case scenario, where $a, o = O(n)$, this simplifies to $O(n^3 + n^2 \cdot label\_num)$. It is important to note that the actual runtime will be significantly influenced by the specific characteristics of the input data, such as the average size of aspect and opinion spans and the distribution of labels within the matrix P. While the worst-case time complexity of Algorithm

1 may appear high, practical considerations such as the typically small size of aspect and opinion spans relative to the entire sentence, and the finite and manageable number of labels, are likely to result in much more efficient execution in real-world scenarios.

3.2.7. Loss function

The design and optimization of loss functions play a pivotal role in guiding the learning process of models towards achieving desired outcomes. This section presents a comprehensive analysis of the loss function employed in our model, which is meticulously crafted to model various relationships between token pairs within a sentence. The overarching goal is to minimize the composite loss function, which is articulated as follows:

$$\mathcal{L} = \mathcal{L}_p + \alpha \left(\mathcal{L}_{psc} + \mathcal{L}_{dep} + \mathcal{L}_{rpd}\right) + \beta \left(\mathcal{L}_{ctd} + \mathcal{L}_{ls}\right) + \lambda \mathcal{L}_{ba} \tag{14}$$

$$\mathcal{L}_p = -\sum_{i=0}^{n+1} \sum_{j=0}^{n+1} \sum_{c \in C_1} \mathbb{I}(y_{ij} = c) \log(p_{ij|c}) \tag{15}$$

The loss function $\mathcal{L}$ is a weighted sum of several component losses, each corresponding to a specific module within the model. These modules are designed to capture distinct aspects of the relationships between token pairs, thereby facilitating a nuanced understanding of the textual data. The primary loss component, $\mathcal{L}_p$, is derived from the cross-entropy function, which quantifies the discrepancy between the predicted and actual labels for token pairs across the entire sentence. This component serves as the foundation of our loss function, ensuring that the model accurately captures the fundamental relationships between tokens.

To further refine the model's performance and enable it to leverage the rich semantic and syntactic information embedded in the text, we introduce additional loss components associated with specific modules, such as $\mathcal{L}_{psc}$, $\mathcal{L}_{dep}$, $\mathcal{L}_{rpd}$, $\mathcal{L}_{ctd}$, and $\mathcal{L}_{ls}$. Each of these components targets a particular aspect of the token pair relationships, ranging from positional semantics to dependency structures and beyond. The Enhanced Label Semantics module, exemplified by $\mathcal{L}_{ls}$, imposes constraints on its sub-modules to accentuate the attention of different modules on distinct token pair relationships:

$$\mathcal{L}_{ls} = -\sum_{i=0}^{n+1} \sum_{j=0}^{n+1} \sum_{c \in C_2} \mathbb{I}(y_{ij} = c) \log(r^{ls}_{ij|c}) \tag{16}$$

The coefficients α, β, and λ are strategically employed to modulate the influence of the respective sub-modules, allowing for a balanced integration of the diverse linguistic features captured by the model. The indicator function $\mathbb{I}(\cdot)$ serves to conditionally activate the loss computation based on the presence of specific labels, thereby ensuring that each component of the loss function is precisely aligned with the model's learning objectives.

The sets of labels C_1 and C_2 delineate the scope of labels applicable to different modules, with C_1 encompassing all labels except for the **S** and **C** labels, as detailed in Table 1. This nuanced approach to label selection underscores the model's capacity to differentiate between various types of token pair relationships, thereby enhancing its ability to extract meaningful insights from textual data.

3.3. Model training

The overall procedure of LSEMH-GCN can be given in Algorithm 2, detailing the step-by-step process of model training and quadruplet extraction. The procedure begins with an input sentence set S with its corresponding truth quadruplets set T and label set L_S. The goal is to output Trained model and a set of sentiment quadruplets, $Quads$, accurately representing the sentiment structure within the sentence.

Step 1: Encoding. The encoder function processes the input sentence S using BERT, generating hidden feature representations H for the sentence and L for the label set. It also computes initial implicit representations for aspects (I_A) and opinions (I_O) using a Multi-Layer Perceptron (MLP) applied to the $[CLS]$ token representation. The output X combines these representations, setting the stage for further processing.

Step 2: Label Semantics. The LabelSemantics function applies Biaffine attention to pairs of tokens from X and the label features L, producing a matrix R^{ls} that captures the semantic relationships between token pairs and labels.

Step 3: Linguistic Features. This step extracts linguistic features from the sentence S, including part-of-speech tags, dependency relations, constituent tree structures, and relative distances between tokens, using tools like Supar and Stanza. These features are represented as matrices R^{psc}, R^{dep}, R^{ctd}, and R^{rpd}.

Step 4: Biaffine Attention. The Biaffine function computes relationships between all token pairs in X, resulting in a matrix R^{ba} that enhances the model's understanding of token interactions.

Step 5: GCN Integration. The Concat function merges all feature matrices, and the MultiLayerGCN function applies graph convolutional operations on X using the combined feature matrix R^{all}, yielding an enriched representation $\overline{X}$.

Step 6: Quadruplet Extraction. The final steps involve concatenating feature representations and applying a prediction function to extract sentiment quadruplets. The VectorConcatenation function integrates $\overline{X}$, R, and R^{ls} into a comprehensive feature vector Z. The GetPrediction function then applies a softmax layer over an MLP to predict sentiment quadruplets, which are extracted and returned as $Quads$.

This training algorithm meticulously integrates various NLP techniques and neural network architectures to achieve precise sentiment quadruplet extraction, demonstrating the model's capability to understand and analyze complex sentence structures and sentiment expressions.

4. Experiments

4.1. Datasets

This section provides a detailed overview of the datasets utilized in our study, which aims to evaluate the efficacy of our proposed method. Our investigation leverages two benchmark datasets, Restaurant-ACOS and Laptop-ACOS, which were meticulously curated by Cai et al. (2021) based on the foundational SemEval 2016 dataset (Pontiki et al., 2016). To ensure a comprehensive evaluation, we adopted the experimental setup delineated by Cai et al. (2021), which involves partitioning each dataset into distinct subsets for training, validation, and testing purposes. This division enables a systematic exploration of the model's performance across different phases of the learning process, ensuring that the results are both robust and generalizable.

Table 2 presents a summary of the statistical characteristics of the Restaurant-ACOS and Laptop-ACOS datasets. The table is structured to provide insights into the composition of the datasets, including the total number of sentences, the distribution of quadruplets across various categories (explicit aspect terms, explicit opinion terms, implicit aspect terms, and implicit opinion terms), and the average number of quadruplets per sentence. Additionally, the table highlights the diversity of categories present in each dataset, underscoring the complexity and richness of the data.

4.2. Implementation details

This section delineates the specific configurations and methodologies employed in the execution of our experiments. Our experimental framework is anchored by the utilization of the Bert_base_uncased model, which serves as the sentence encoder. This choice is motivated by BERT's proven efficacy in capturing deep contextualized representations of textual data. To facilitate the optimization process, we employed the AdamW optimizer (Loshchilov & Hutter, 2017),

Table 2
Summary of experimental dataset statistics: The dataset is categorized into four types of terms: 'EA' for explicit aspect terms, 'EO' for explicit opinion terms, 'IA' for implicit aspect terms, and 'IO' for implicit opinion terms. Each category is quantitatively represented to facilitate comprehensive analysis.

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

Algorithm 2 Training of the LSEMH-GCN model

Require: Sentence set S, set of truth quadruplets from the sentence set T and label set L_S.
Ensure: Trained model LSEMH-GCN and set of quadruplets predicted by the model $Quads$.

repeat
  train(LSEMH-GCN(S, L_S), T)
until Convergence
$Quads \leftarrow$ LSEMH-GCN(S, L_S)
return LSEMH-GCN, $Quads$
function LSEMH-GCN(S, L_S)
  $X, L \leftarrow$ Encoder(S, L_S)
  $R^{ls} \leftarrow$ LabelSemantics(X, L)
  $R^{psc}, R^{dep}, R^{ctd}, R^{rpd} \leftarrow$ LinguisticFeatures(S)
  $R^{ba} \leftarrow$ Biaffine(X, X)
  $R^{all} \leftarrow$ Concat($R^{psc}, R^{dep}, R^{ctd}, R^{rpd}, R^{ba}, R^{ls}$)
  $\overline{X} \leftarrow$ MultiLayerGCN(X, R^{all})
  $R \leftarrow$ Concat($R^{psc}, R^{dep}, R^{ctd}, R^{rpd}, R^{ba}$)
  $Z \leftarrow$ VectorConcatenation($\overline{X}, R, R^{ls}$)
  $P \leftarrow$ GetPrediction(Z)
  $Quads \leftarrow$ QuadrupletExtraction(P, len(l_S))
  return $Quads$
end function
function Encoder(S, L_S)
  $H \leftarrow$ Bert(S)
  $L \leftarrow$ Bert(L_S)
  $I_A \leftarrow$ MLP($H[CLS]$)
  $I_O \leftarrow$ MLP($H[CLS]$)
  $X \leftarrow I_A + H[1:] + I_O$
  return X, L
end function
function LabelSemantics(X, L)
  $Y \leftarrow$ GetTokenPair(X)
  $R^{ls} \leftarrow$ Biaffine(Y, L)
  return R^{ls}
end function
function LinguisticFeatures(S)
  $R^{ctd} \leftarrow$ Supar(S)
  $R^{psc}, R^{dep} \leftarrow$ Stanza(S)
  $R^{rpd} \leftarrow$ GetRelativeDistance(S)
  return $R^{psc}, R^{dep}, R^{ctd}, R^{rpd}$
end function
function GetPrediction(Z)
  return SoftMax(MLP(Z))
end function

renowned for its effectiveness in handling sparse gradients and its incorporation of weight decay to prevent overfitting.

The fine-tuning of the BERT model was conducted with a learning rate of 10^{-6}, a setting that strikes a balance between adaptation and preservation of the pre-trained weights. For other trainable parameters within our model, a slightly higher learning rate of 10^{-5} was adopted to expedite the convergence of the training process. To mitigate the risk of overfitting, a dropout rate of 0.5 was implemented, serving as a regularization technique by randomly omitting a subset of features during each iteration of training.

The architecture of our model comprises hidden state dimensions of 768 for BERT and 400 for the GCN, reflecting a tailored approach to accommodate the distinct characteristics of each module. The LSEMH-GCN model was subjected to 150 epochs of iterative training, a duration determined to be sufficient for the model to converge and exhibit stable performance.

Dataset-specific configurations were also meticulously calibrated to optimize the training process. For the Restaurant-ACOS dataset, a batch size of 21 was selected, while for the Laptop-ACOS dataset, a smaller batch size of 5 was deemed more appropriate due to its distinct characteristics. The parameters α, β, and λ were set to 0.01, 0.05, and 0.1, respectively, to finely control the influence of each module within the model.

Model performance was rigorously evaluated based on precision, recall, and F1 score, metrics that collectively provide a holistic view of the model's ability to accurately and comprehensively extract aspect-category opinion sentiment quadruplets. A quadruplet was deemed correct only if all four elements, along with their combinations, precisely matched those of the target quadruplet.

Linguistic features, including part-of-speech tags and dependency relations, were generated using Stanza (Qi et al., 2020), a state-of-the-art natural language processing toolkit. Additionally, the constituent tree was derived from the SuPar parser, further enriching the model's input with syntactic information.

4.3. Baselines

In the baseline section, we selected a series of state-of-the-art models from different time periods as baseline models for comparison with the LSEMH-GCN model. The rationale behind choosing these baseline models is that they represent various methods proposed in the development of the ASQE field, encompassing a broad spectrum of approaches ranging from pipeline and end-to-end methods to generation-based and machine reading comprehension strategies. This selection was made to ensure a comprehensive evaluation of the LSEMH-GCN model's performance and to reveal its advantages relative to existing methods in the field. To delve deeper into these methods, these baseline models were categorized into four groups based on their foundational approaches to facilitate a more effective comparison between them and our proposed approach.

Pipeline Methods:

1. **Double Propagation (DP)** Qiu et al. (2011) pioneers a rule-based strategy for the extraction of aspect-opinion-sentiment (AOS) triplets. By adeptly leveraging syntactic relations, DP meticulously extracts AOS triplets and judiciously assigns sentiment and aspect categories using sentiment dictionaries. This method has significantly advanced the understanding of how syntactic structures can be harnessed for sentiment analysis.
2. **Extract-Classify (EC)** Cai et al. (2021) innovatively undertakes the joint extraction of aspect-opinion pairs, seamlessly followed by the prediction of category-sentiment for each pair. This integration of extraction and classification tasks has contributed

Table 3
The experimental results of all baselines on the two benchmark datasets.

METHOD	Restaurant-ACOS			Laptop-ACOS		
	P	R	F1	P	R	F1
DP	34.67	15.08	21.04	13.04	0.57	8.00
Extract-Classify	38.54	52.96	44.61	45.56	29.48	35.80
JET	59.81	28.94	39.01	44.52	16.25	23.81
TAS-BERT	26.29	46.29	33.53	47.15	19.22	27.31
BARTABSA	56.62	55.35	55.98	41.65	40.46	41.05
GAS	60.69	58.52	59.59	41.60	42.75	42.17
Paraphrase	58.98	59.11	59.04	41.77	45.04	43.34
Opinion tree generation	63.96	61.74	62.83	46.11	44.79	45.44
EMRC	64.97	61.18	63.02	47.27	44.66	45.92
Our LSEMH-GCN	**66.00**	**62.61**	**64.26**	**48.81**	**45.43**	**47.06**

to the refinement of joint modeling approaches in sentiment analysis.

3. **JET** Xu et al. (2020) introduces an end-to-end framework for the extraction of AOS triplets, incorporating a position-aware labeling scheme. Adapted for the ASQE task, JET's methodology of first extracting triplets before predicting their corresponding aspect categories has enriched the landscape of end-to-end sentiment analysis models.

End-to-End Method:

4. **TAS-BERT** Wan et al. (2020) adeptly executes the joint extraction of aspect-opinion pairs under category-sentiment conditions, employing an input transformation strategy and a filtering mechanism to derive quadruplets. This approach has been instrumental in demonstrating the efficacy of transformer-based models for complex sentiment analysis tasks.

Generation-Based Methods:

5. **BARTABSA** Yan et al. (2021) ingeniously reformulates all ABSA subtasks into a unified generation task, with a focus on generating category indices. This model has been pivotal in showcasing the versatility and potential of generative models in sentiment analysis.
6. **GAS** Zhang et al. (2021b) proposes a unified generation framework for ABSA tasks, conceptualizing them as sentiment element sequence generation challenges. This innovative approach has broadened the applicability of generation models in sentiment analysis.
7. **Paraphrase** Zhang et al. (2021a) introduces a paraphrasing modeling paradigm for the joint detection of sentiment elements in quadruplets, transforming ABSA tasks into paraphrase generation processes. This method has contributed to the exploration of novel paradigms in sentiment analysis.
8. **Opinion Tree Generation** Bao et al. (2022) envisions a tree-structured semantic representation for the joint detection of sentiment elements, offering a nuanced representation of sentiment structures.

Machine Reading Comprehension Method:

9. **EMRC** Ye et al. (2023) adopts a multi-turn machine reading comprehension (MRC) approach to the ASQE task, showcasing the versatility of MRC methodologies in handling complex extraction tasks. This method has underscored the potential of adapting MRC techniques for sentiment analysis.

Within these baseline models, the first three models utilize traditional pipeline methods, and comparing them clearly reveals the performance advantages of end-to-end methods. It is particularly noteworthy that both the Double Propagation (DP) (Qiu et al., 2011) and JET (Xu et al., 2020) models start from excellent sentiment triplet models and extract quadruplets by incorporating a triplet category classification structure. This highlights the difficulty of the ASQE task compared to models specifically designed for quadruplet extraction, thereby demonstrating the necessity and value of designing models specifically for ASQE tasks. The subsequent six baseline models avoid potential error propagation issues in pipeline models and exhibit higher performance. In particular, the generation-based and machine reading comprehension methods deeply deconstruct the ASQE task, enhancing the fusion among ASQE subtasks and achieving better results. However, these methods do not fully leverage graph convolutional networks and language analysis techniques. In contrast, the LSEMH-GCN model is an end-to-end model based on graph convolutional networks. It cleverly integrates language analysis techniques and adopts an optimized structure specifically designed for ASQE tasks. Such design decisions have resulted in higher performance in experiments, fully demonstrating the effectiveness of combining graph convolutional networks with language analysis techniques. These results further highlight the advantages of the LSEMH-GCN model in addressing ASQE tasks.

4.4. Main results

According to the main experimental results, our LSEMH-GCN model shows outstanding performance in terms of the F1 metric, as shown in Table 3. Bold results indicate the best performance, and underlined results indicate the second-best performance. Compared to pipeline methods, end-to-end methods, and MRC-based methods, our model achieves better performance on both datasets. It is observed that end-to-end and MRC-based methods exhibit more significant improvements compared to pipeline methods. This is because they can establish correlations between subtasks and alleviate error propagation issues by jointly training multiple subtasks. Compared to the baseline models, our LSEMH-GCN model achieves an average F1 score improvement of 15.52% and 12.30% on the Restaurant-ACOS and Laptop-ACOS datasets, respectively. This improvement can be attributed to our model's ability to learn word representations by leveraging relationships between words and linguistic knowledge, as well as the enhanced category classification through label semantic enhancement.

4.5. Model analysis

4.5.1. Ablation study

The LSEMH-GCN model incorporates several innovative modules designed to enhance its performance in extracting aspect-category-opinion-sentiment quadruplets. To rigorously evaluate the contribution of these modules, we systematically removed each one in turn and observed the resultant impact on the model's performance. The outcomes and score comparisons of these experiments are documented in Table 4, Fig. 6, and Fig. 7, respectively, facilitating a nuanced understanding of the role each module plays.

w/o Asymmetric Vector Concatenation: The removal of the asymmetric vector concatenation strategy, pivotal for enabling the model to discern between the upper and lower triangular regions of the label matrix, results in a decrement in performance metrics. Specifically, the precision, recall, and F1 scores on the Restaurant-ACOS dataset slightly

Table 4
Performance of ablation experiments on the two datasets.

METHOD	Restaurant-ACOS			Laptop-ACOS		
	P	R	F1	P	R	F1
LSEMH-GCN	66.00	62.61	64.26	48.81	45.43	47.06
w/o Asymmetric Vector Concatenation	65.81	60.21	62.89	52.67	39.70	45.27
w/o Label Semantics	64.10	61.72	62.89	47.05	42.60	44.71
w/o Constituent Tree-based Distance	65.58	60.90	63.15	46.60	41.10	43.68

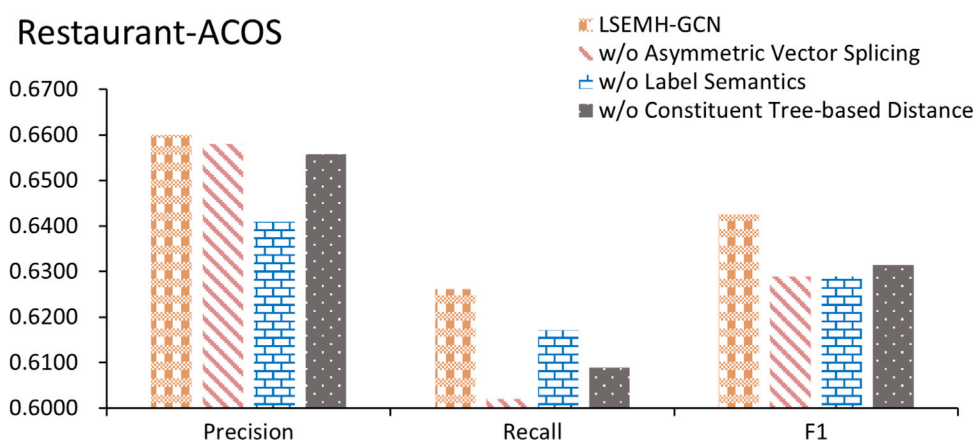

Fig. 6. The results of the ablation study on the dataset Restaurant-ACOS.

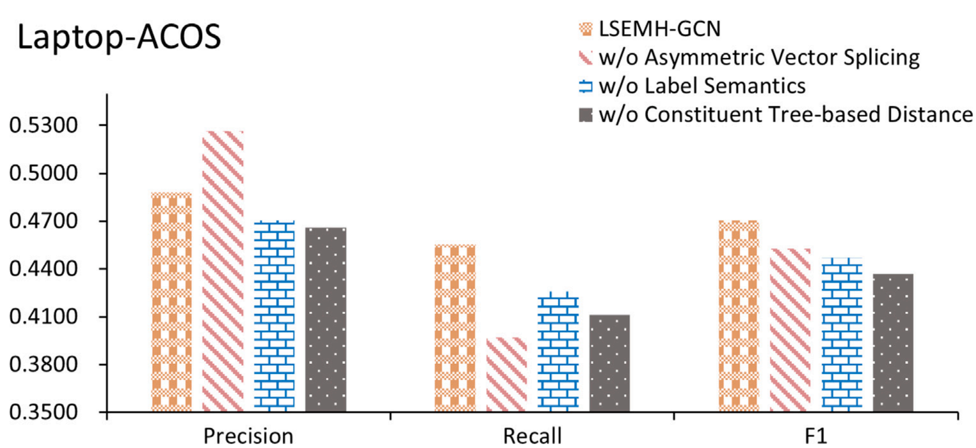

Fig. 7. The results of the ablation study on the dataset Laptop-ACOS.

decrease to 65.81, 60.21, and 62.89, respectively, from the original scores of 66.00, 62.61, and 64.26. This decline is more pronounced on the Laptop-ACOS dataset, where precision increases to 52.67 due to the model's inability to effectively differentiate between category classification and sentiment polarity, essential for accurate quadruplet formation, while recall and F1 scores significantly drop to 39.70 and 45.27, respectively.

w/o Label Semantics: Excluding the label semantics module, responsible for aggregating token information across sentiment polarity and category classification dimensions, leads to a reduction in the model's performance. This is evidenced by the decrease in precision, recall, and F1 scores to 64.10, 61.72, and 62.89, respectively, on the Restaurant-ACOS dataset, and to 47.05, 42.60, and 44.71 on the Laptop-ACOS dataset. This underscores the critical role of label semantics in enhancing the model's comprehension and processing capabilities.

w/o Constituent Tree-based Distance: Substituting the constituent tree-based relative distance with a syntax dependency tree-based relative distance, which may introduce noise, notably impacts the model's performance, especially on the Laptop-ACOS dataset. This is reflected in the precision, recall, and F1 scores, which adjust to 65.58, 60.90, and 63.15 for the Restaurant-ACOS dataset, and more significantly to 46.60, 41.10, and 43.68 for the Laptop-ACOS dataset. This dataset's broader classification categories amplify the adverse effects of this modification, validating the original choice of constituent tree-based distance.

The results of the ablation study unequivocally affirm the rationality and effectiveness of the modular design within the LSEMH-GCN model. Each module's contribution to the model's performance can clearly articulated.

4.5.2. Token pair label prediction via label semantics module

The visualization, as depicted in Fig. 8, employs a three-dimensional scatter plot to articulate the nuanced interactions between tokens and their corresponding labels as interpreted by the Label Semantics Module. The axes of this plot are defined as follows: the x and y axes delineate the tokens under consideration, while the z axis enumerates the spectrum of possible labels. Within this schema, specific z values are assigned semantic significance; for instance, $z = 0$ is designated as the "Nonesense Label", indicating a classification outside the bounds of sentiment or category labels. Conversely, $z = 3$

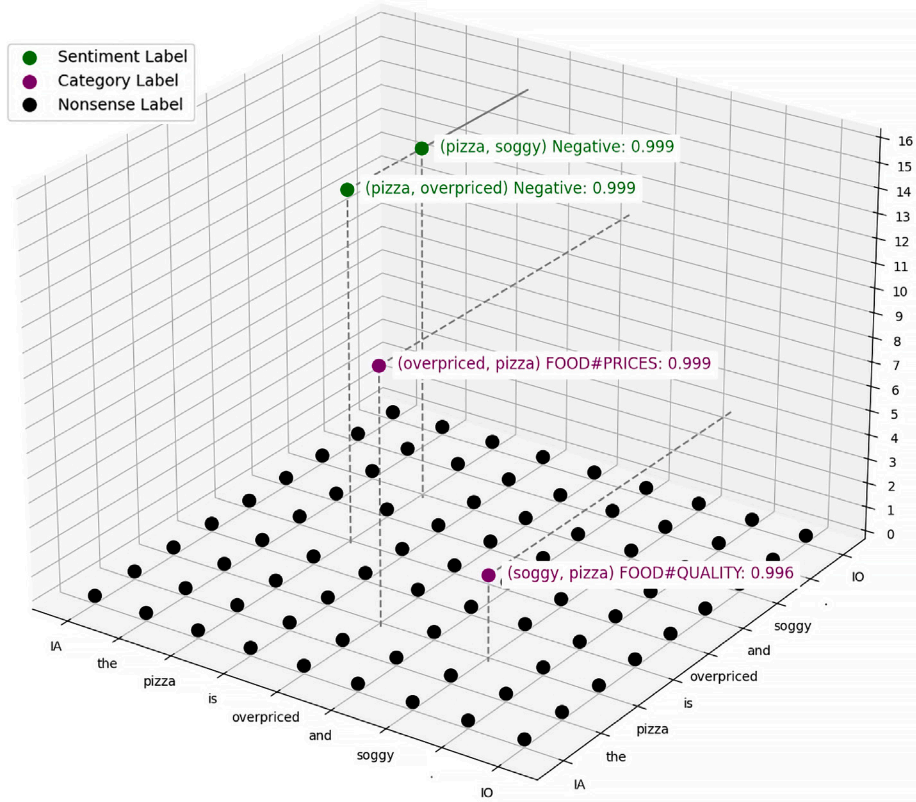

Fig. 8. Visualizations of token pair label predictions from the label semantics module on the sentence "the pizza is overpriced and soggy".

is attributed to the "FOOD#QUALITY" category label, and $z = 14$ is reserved for the "Negative" sentiment label. Each plotted data point within this three-dimensional space represents the Label Semantics Module's prediction for a given token-pair, annotated in the format $(token_1,\ token_2)Label\ :\ Probability$.

A critical observation from this visualization is the high degree of accuracy with which the Label Semantics Module operates. The probability of correct label assignment for each data point prominently exceeds the 90% threshold, underscoring the module's proficiency in accurately predicting labels for token-pairs. This visualization not only serves as a testament to the module's effectiveness but also provides valuable insights into the underlying mechanisms that facilitate such precise label prediction.

4.5.3. Investigating the impact of constituent vs. Dependency tree distances on token representations of linguistic features module

This section delves into an experimental analysis aimed at elucidating the impact of utilizing constituent versus dependency tree distances on the representation of tokens within the context of a GCN as part of the linguistic features module. The crux of this investigation lies in addressing the challenge posed by syntactic dependency trees, where tokens pertaining to different aspects may exhibit relatively small distances, potentially leading to mutual interference during the feature aggregation process. This interference can result in overly similar feature representations for tokens of distinct aspects, thereby diminishing the model's ability to accurately discern and process varying sentiment polarities within a sentence.

To rigorously explore the differential effects of constituent and dependency tree-based relative distances on the aggregation efficacy of the Linguistic Features Module, we embarked on a methodical analysis. This involved obtaining token representation vectors for aspect and sentiment words within a sentence, post-aggregation by the Linguistic Features Module, incorporating either dependency or constituent tree-based distances. Subsequently, we computed the Manhattan distance between these token representation vectors in their original high-dimensional space. To facilitate a more intuitive understanding, we projected these vectors onto a two-dimensional plane using Principal Component Analysis (PCA).

The visualization of these projections is presented in Figs. 9 and 10, corresponding to token representation vectors aggregated with constituent and dependency tree relative distances, respectively. In these figures, blue data points denote the mapping positions of token representation vectors within the two-dimensional space, with "Manhattan D" signifying the Manhattan distance between two tokens in the original high-dimensional space. A notable observation from Fig. 9 is the discernible expansion in the distance between the aspect word "food" and the sentiment words "upgraded" and "amazing", as well as between the aspect word "decor" and these sentiment words, when compared to the distances observed in Fig. 10. Specifically, the distance differential for "food" and its associated sentiment words is augmented by 19.6%, and for "decor" by 20.8%, in the constituent tree-based representation relative to the dependency tree-based representation. This expansion in distance underscores a reduced similarity between tokens from different aspects in the constituent tree-based vector space, thereby enhancing the model's capacity to distinguish between disparate aspect information within a sentence. Moreover, these findings affirm the superiority of constituent tree-based relative distances in mitigating the interference issues commonly associated with dependency tree-based distances. Through this experimental analysis, we not only illuminate the critical role of tree-based relative distances in shaping token representations but also underscore the efficacy of constituent trees in fostering more distinct and interpretable feature representations.

4.5.4. Case study

This section presents a detailed case study aimed at elucidating the impact of specific vector concatenation methods on the predictive

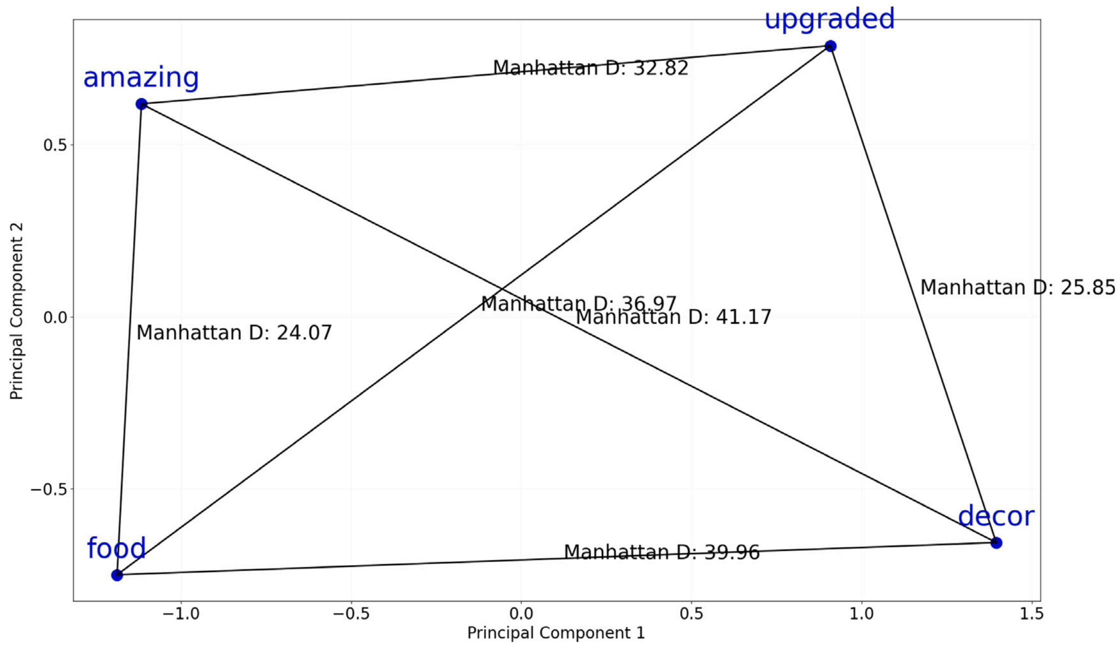

Fig. 9. PCA visualization of token representations from the linguistic features module aggregated using constituent tree relative distances.

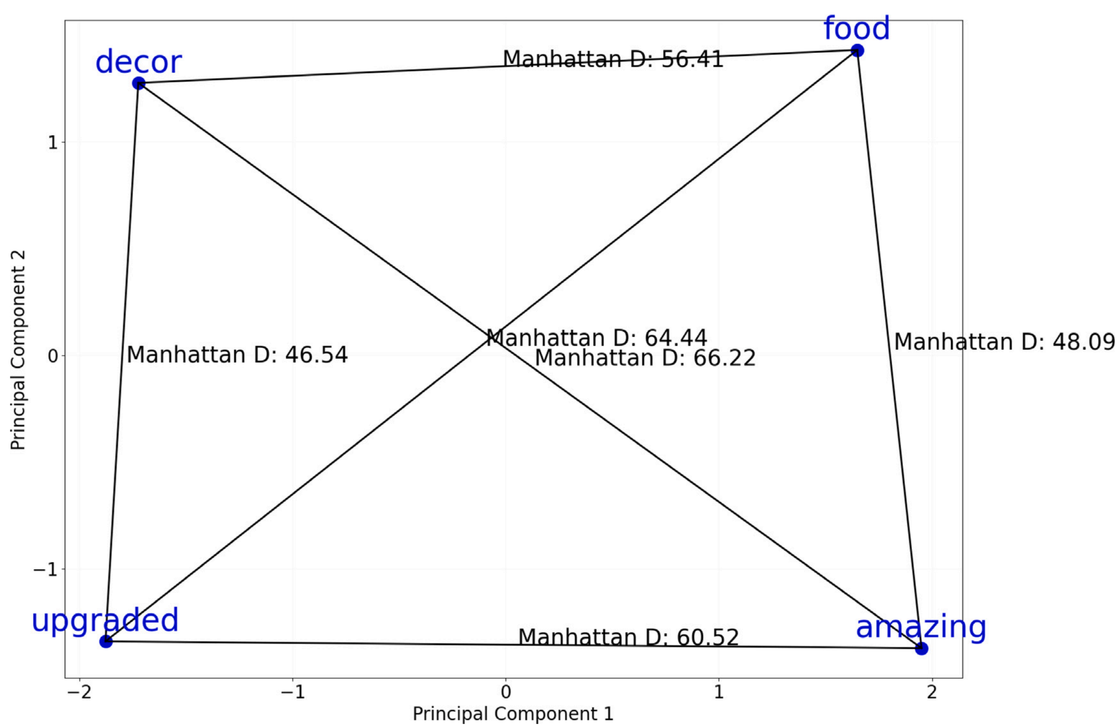

Fig. 10. PCA visualization of token representations from the linguistic features module aggregated using dependency tree relative distances.

capabilities of a given model. Through the examination of two distinct reviews, we aim to provide a comprehensive analysis of the influence exerted by the implementation of specific vector concatenation methods on the model's ability to predicate accurate sentiment quadruplets. The predicted results and corresponding visualizations for both reviews, with and without the application of specific vector concatenation methods, are depicted in Fig. 11.

In the Review 1, the case "the fish was really, really fresh." presents an interesting scenario. When the LSEMH-GCN model, equipped with asymmetric vector concatenation, processes this case, it successfully predicts the sentiment quadruplet (FOOD#QUALITY, fish, fresh, positive). In contrast, the absence of asymmetric vector concatenation (w/o) leads to the absence of the aspect category (FOOD#QUALITY), thereby yielding an incomplete prediction. This observation underscores the critical role of specific vector concatenation in preserving the integrity of sentiment quadruplets.

In the Review 2, the case "quick and friendly service." further corroborates the significance of specific vector concatenation methods. The LSEMH-GCN model, with the implementation of these methods, adeptly identifies two distinct sentiment quadruplets: (SERVICE#GENERAL, service, quick, positive) and (SERVICE#GENERAL, service,

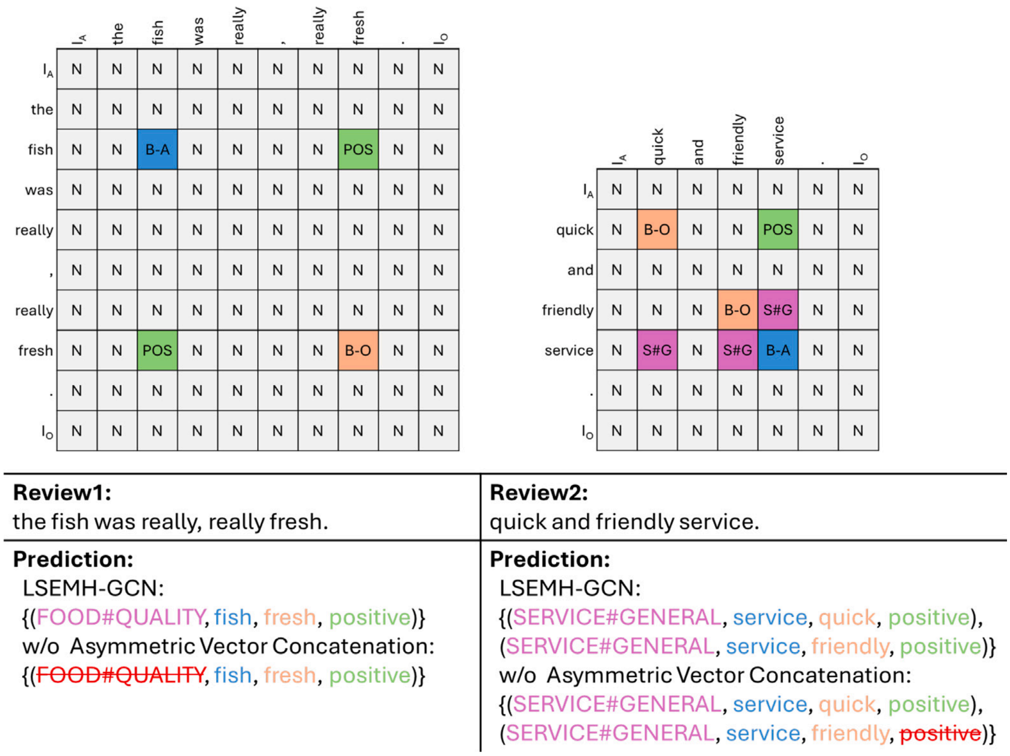

Fig. 11. Illustrating token pair label predictions generated by the label semantic module.

friendly, positive). Remarkably, the model's performance remains consistent even without asymmetric vector concatenation for this particular review, suggesting that the effectiveness of specific vector concatenation methods may vary depending on the linguistic and contextual nuances of the input text.

It is observed that the omission of specific vector concatenation strategies often results in incomplete predictions of sentiment quadruplets. This inadequacy primarily stems from the model's inability to fully extract the entire quadruplet, a consequence of the feature vectors becoming overly similar post-aggregation of aspect and opinion terms through the GCN. Furthermore, the symmetric nature of linguistic features exacerbates this challenge, rendering the model incapable of effectively distinguishing between vectors representative of category classification and sentiment polarity. This limitation also impedes the model's ability to differentiate between the upper and lower triangular regions of the matrix, potentially leading to repeated predictions of sentiment polarity or category classification and, consequently, to the omission of essential elements of the quadruplet.

In contrast, the adoption of specific vector concatenation methods markedly enhances the model's predictive accuracy. By enabling the model to distinctly differentiate between category vectors and sentiment vectors, specific vector concatenation methods facilitate the complete and accurate prediction of all elements constituting the sentiment quadruplet. This approach not only addresses the challenges posed by the similarity of feature vectors and the symmetric nature of linguistic features but also significantly improves the model's ability to navigate the complexities of sentiment analysis.

4.5.5. Conclusion and future work

This paper has introduced the LSEMH-GCN architecture, a novel and sophisticated framework specifically designed to navigate the intricacies of the ASQE task. The LSEMH-GCN architecture represents an improvement forward in sentiment analysis, characterized by its innovative integration of multiple linguistic features and its strategic focus on semantic information. The architecture's ability to establish meaningful structured representations between tokens and their implicit counterparts, coupled with the implementation of a multi-layer heterogeneous graph convolutional network, has markedly enhanced its interpretability and performance. A key contribution of our model is the specific vector concatenation method, which has proven instrumental in enabling precise differentiation between category and sentiment vectors, thereby facilitating accurate sentiment analysis. Furthermore, the model's unique asymmetric labeling scheme for the extraction of sentiment quadruplets underscores its advanced functionalities and potential to redefine the sentiment analysis landscape.

The empirical validation of the LSEMH-GCN model, through extensive experimental evaluations on benchmark datasets, has demonstrated its superior performance, consistently outperforming all baseline models across various metrics. This underscores the effectiveness and robustness of the LSEMH-GCN architecture.

Our future endeavors will focus on the exploration of hierarchical category classification methods. This initiative is aimed at further enhancing the model's classification efficiency by capitalizing on the inherent structure of category labels. By adopting a more nuanced approach to category classification, we are optimistic about achieving even greater levels of accuracy and efficiency in sentiment analysis tasks.

CRediT authorship contribution statement

Yiheng Fu: Conceptualization, Data curation, Formal analysis, Software, Visualization, Writing – original draft. **Xiaoliang Chen:** Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. **Duoqian Miao:** Funding acquisition. **Xiaolin Qin:** Funding acquisition, Resources. **Peng Lu:** Writing – review & editing. **Xianyong Li:** Validation.

Declaration of competing interest

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

Data availability

To support the reproducibility of our findings and encourage further research, this paper provides comprehensive access to the complete program code, datasets, and detailed instructions. This transparency ensures that interested researchers can replicate the experiments and potentially extend the work presented herein. Code and data are available for download at the following web links https://github.com/Thirring/LSEMH-GCN.

Acknowledgments

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

References

Bao, X., Wang, Z., Jiang, X., Xiao, R., & Li, S. (2022). Aspect-based sentiment analysis with opinion tree generation. In L. D. Raedt (Ed.), *Proceedings of the thirty-first international joint conference on artificial intelligence, IJCAI 2022, vienna, Austria, 23-29 July 2022* (pp. 4044–4050). ijcai.org, http://dx.doi.org/10.24963/ijcai.2022/561.

Cai, H., Xia, R., & Yu, J. (2021). Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 1: long papers), virtual event, August 1-6, 2021* (pp. 340–350). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.29.

Chen, S., Liu, J., Wang, Y., Zhang, W., & Chi, Z. (2020). Synchronous double-channel recurrent network for aspect-opinion pair extraction. In D. Jurafsky, J. Chai, N. Schluter, & J. R. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics, ACL 2020, online, July 5-10, 2020* (pp. 6515–6524). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-main.582.

Chen, Y., Tang, Y., & Yuan, Y. (2023). Attention-enhanced graph neural networks with global context for session-based recommendation. *IEEE Access*, *11*, 26237–26246. http://dx.doi.org/10.1109/ACCESS.2023.3254897.

Chen, C., Teng, Z., & Zhang, Y. (2020). Inducing target-specific latent structures for aspect sentiment classification. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing, EMNLP 2020, online, November 16-20, 2020* (pp. 5596–5607). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.emnlp-main.451.

Chen, H., Zhai, Z., Feng, F., Li, R., & Wang, X. (2022). Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers), ACL 2022, Dublin, Ireland, May 22-27, 2022* (pp. 2974–2985). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2022.acl-long.212.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, volume 1 (long and short papers)* (pp. 4171–4186). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/n19-1423.

Dozat, T., & Manning, C. D. (2017). Deep biaffine attention for neural dependency parsing. In *5th international conference on learning representations, ICLR 2017, Toulon, France, April 24-26, 2017, conference track proceedings*. OpenReview.net, https://openreview.net/forum?id=Hk95PK9le.

Fan, Z., Wu, Z., Dai, X., Huang, S., & Chen, J. (2019). Target-oriented opinion words extraction with target-fused neural sequence labeling. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, volume 1 (long and short papers)* (pp. 2509–2518). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/n19-1259.

Gou, Z., Guo, Q., & Yang, Y. (2023). Mvp: Multi-view prompting improves aspect sentiment tuple prediction. In A. Rogers, J. L. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: long papers), ACL 2023, Toronto, Canada, July 9-14, 2023* (pp. 4380–4397). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2023.acl-long.240.

Guo, Z., Zhang, Y., & Lu, W. (2019). Attention guided graph convolutional networks for relation extraction. In A. Korhonen, D. R. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th conference of the association for computational linguistics, ACL 2019, florence, Italy, July 28- August 2, 2019, volume 1: long papers* (pp. 241–251). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/p19-1024.

Hu, M., Zhao, S., Guo, H., Xue, C., Gao, H., Gao, T., Cheng, R., & Su, Z. (2021). Multi-label few-shot learning for aspect category detection. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 1: long papers), virtual event, August 1-6, 2021* (pp. 6330–6340). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.495.

Li, R., Chen, H., Feng, F., Ma, Z., Wang, X., & Hovy, E. H. (2021). Dual graph convolutional networks for aspect-based sentiment analysis. In C. Zong, F. Xia, W. Li & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 1: long papers), virtual event, August 1-6, 2021* (pp. 6319–6329). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.494.

Liang, S., Wei, W., Mao, X., Wang, F., & He, Z. (2022). Bisyn-gat+: Bi-syntax aware graph attention network for aspect-based sentiment analysis. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Findings of the association for computational linguistics: ACL 2022, Dublin, Ireland, May 22-27, 2022* (pp. 1835–1848). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2022.findings-acl.144.

Loshchilov, I., & Hutter, F. (2017). Fixing weight decay regularization in adam. *CoRR*, arXiv:1711.05101.

Lu, Q., Sun, X., Sutcliffe, R. F. E., Xing, Y., & Zhang, H. (2022). Sentiment interaction and multi-graph perception with graph convolutional networks for aspect-based sentiment analysis. *Knowledge-Based Systems*, *256*, Article 109840. http://dx.doi.org/10.1016/j.knosys.2022.109840.

Ma, D., Li, S., Wu, F., Xie, X., & Wang, H. (2019). Exploring sequence-to-sequence learning in aspect term extraction. In A. Korhonen, D. R. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th conference of the association for computational linguistics, ACL 2019, florence, Italy, July 28- August 2, 2019, volume 1: long papers* (pp. 3538–3547). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/p19-1344.

Ma, D., Li, S., Zhang, X., & Wang, H. (2017). Interactive attention networks for aspect-level sentiment classification. In C. Sierra (Ed.), *Proceedings of the twenty-sixth international joint conference on artificial intelligence, IJCAI 2017, melbourne, Australia, August 19-25, 2017* (pp. 4068–4074). ijcai.org, http://dx.doi.org/10.24963/ijcai.2017/568.

Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., Al-Ayyoub, M., Zhao, Y., Qin, B., Clercq, O. D., Hoste, V., Apidianaki, M., Tannier, X., Loukachevitch, N. V., Kotelnikov, E. V., Bel, N., Zafra, S. M. J., & Eryigit, G. (2016). SemEval-2016 task 5: Aspect based sentiment analysis. In S. Bethard, D. M. Cer, M. Carpuat, D. Jurgens, P. Nakov, & T. Zesch (Eds.), *Proceedings of the 10th international workshop on semantic evaluation, semEval@nAACL-HLT 2016, san diego, CA, USA, June 16-17, 2016* (pp. 19–30). The Association for Computer Linguistics, http://dx.doi.org/10.18653/v1/s16-1002.

Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A python natural language processing toolkit for many human languages. In A. Celikyilmaz, & T. Wen (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations, ACL 2020, online, July 5-10, 2020* (pp. 101–108). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.acl-demos.14.

Qiu, M., Huang, X., Chen, C., Ji, F., Qu, C., Wei, W., Huang, J., & Zhang, Y. (2021). Reinforced history backtracking for conversational question answering. In *Thirty-fifth AAAI conference on artificial intelligence, AAAI 2021, thirty-third conference on innovative applications of artificial intelligence, IAAI 2021, the eleventh symposium on educational advances in artificial intelligence, EAAI 2021, virtual event, February 2-9, 2021* (pp. 13718–13726). AAAI Press, http://dx.doi.org/10.1609/aaai.v35i15.17617.

Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, *37*(1), 9–27. http://dx.doi.org/10.1162/coli_a_00034.

Wan, H., Yang, Y., Du, J., Liu, Y., Qi, K., & Pan, J. Z. (2020). Target-aspect-sentiment joint detection for aspect-based sentiment analysis. In *The thirty-fourth AAAI conference on artificial intelligence, AAAI 2020, the thirty-second innovative applications of artificial intelligence conference, IAAI 2020, the tenth AAAI symposium on educational advances in artificial intelligence, EAAI 2020, New york, NY, USA, February 7-12, 2020* (pp. 9122–9129). AAAI Press, http://dx.doi.org/10.1609/aaai.v34i05.6447.

Wei, W., Liu, J., Mao, X., Guo, G., Zhu, F., Zhou, P., Hu, Y., & Feng, S. (2021). Target-guided emotion-aware chat machine. *ACM Transactions on Information Systems*, *39*(4), 43:1–43:24. http://dx.doi.org/10.1145/3456414.

Xiong, H., Yan, Z., Wu, C., Lu, G., Pang, S., Xue, Y., & Cai, Q. (2023). BART-based contrastive and retrospective network for aspect-category-opinion-sentiment quadruple extraction. *International Journal of Machine Learning and Cybernetics*, *14*(9), 3243–3255. http://dx.doi.org/10.1007/s13042-023-01831-8.

Xu, L., Chia, Y. K., & Bing, L. (2021). Learning span-level interactions for aspect sentiment triplet extraction. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 1: long papers), virtual event, August 1-6, 2021* (pp. 4755–4766). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.367.

Xu, L., Li, H., Lu, W., & Bing, L. (2020). Position-aware tagging for aspect sentiment triplet extraction. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing, EMNLP 2020, online, November 16-20, 2020* (pp. 2339–2349). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2020.emnlp-main.183.

Yan, H., Dai, J., Ji, T., Qiu, X., & Zhang, Z. (2021). A unified generative framework for aspect-based sentiment analysis. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 1: long papers), virtual event, August 1-6, 2021* (pp. 2416–2429). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-long.188.

Ye, S., Zhai, Z., & Li, R. (2023). Enhanced machine reading comprehension method for aspect sentiment quadruplet extraction. In K. Gal, A. Nowé, G. J. Nalepa, R. Fairstein, & R. Radulescu (Eds.), *Frontiers in artificial intelligence and applications: 372, ECAI 2023 - 26th European conference on artificial intelligence, September 30 - October 4, 2023, krakóW, Poland - including 12th conference on prestigious applications of intelligent systems (PAIS 2023)* (pp. 2874–2881). IOS Press, http://dx.doi.org/10.3233/FAIA230600.

Yu, W., Wang, X., Yang, P., Xiao, Y., & Wang, J. (2022). Aspect-level sentiment analysis based on graph attention fusion networks. In *24th IEEE int conf on high performance computing & communications; 8th int conf on data science & systems; 20th int conf on smart city; 8th int conf on dependability in sensor, cloud & big data systems & application, HPCC/DSS/smartCity/dependSys 2022, Hainan, China, December 18-20, 2022* (pp. 1776–1781). IEEE, http://dx.doi.org/10.1109/HPCC-DSS-SmartCity-DependSys57074.2022.00268.

Zhang, W., Deng, Y., Li, X., Yuan, Y., Bing, L., & Lam, W. (2021). Aspect sentiment quad prediction as paraphrase generation. In M. Moens, X. Huang, L. Specia, & S. W. Yih (Eds.), *Proceedings of the 2021 conference on empirical methods in natural language processing, EMNLP 2021, virtual event / punta cana, Dominican Republic, 7-11 November, 2021* (pp. 9209–9219). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.emnlp-main.726.

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2021). Towards generative aspect-based sentiment analysis. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing, ACL/IJCNLP 2021, (volume 2: short papers), virtual event, August 1-6, 2021* (pp. 504–510). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2021.acl-short.64.

Zhou, T., Shen, Y., Chen, K., & Cao, Q. (2023). Hierarchical dual graph convolutional network for aspect-based sentiment analysis. *Knowledge-Based Systems*, *276*, Article 110740. http://dx.doi.org/10.1016/j.knosys.2023.110740.

Zhou, J., Yang, H., He, Y., Mou, H., & Yang, J. (2023). A unified one-step solution for aspect sentiment quad prediction. In A. Rogers, J. L. Boyd-Graber, & N. Okazaki (Eds.), *Findings of the association for computational linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023* (pp. 12249–12265). Association for Computational Linguistics, http://dx.doi.org/10.18653/v1/2023.findings-acl.777.