Expert Systems With Applications 299 (2026) 130088

Contents lists available at ScienceDirect

Expert Systems With Applications

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

DGSEP: Dual-stage generative model with sequence-oriented labeling and element-to-tuple prompting improves aspect sentiment triplet extraction

Yujun Chen[a], Mingwei Tang[a,*], Shangyi Du[b], Kun Yang[a], Yanxi Zheng[a], Mingfeng Zhao[c]

[a] School of Computer and Software Engineering, Xihua University, 610039, Chengdu, China
[b] School of Statistics and Computer Science, McGill University, H3A 0G4, Montreal, Quebec, Canada
[c] China Mobile Group Design Institute Co., Ltd. Sichuan Branch, 610045, Chengdu Sichuan Province, China

ARTICLE INFO

Keywords:
Aspect sentiment triplet extraction
Dual-stage process
Prompt templates
Label-oriented sequence label generation fusion module

ABSTRACT

Generative methods have made significant progress in the Aspect Sentiment Triplet Extraction (ASTE) task and have attracted widespread attention. However, existing studies typically adopt a single-stage generation strategy and do not fully explore the potential of step-by-step generation. In this paper, we propose the DGSEP (Dual-stage generative model with sequence-oriented labeling and element-to-tuple prompting improves aspect sentiment triplet extraction) framework to address this gap. The DGSEP framework adopts a dual-stage process. In the first step, we independently predict each individual element (i.e., aspect, sentiment, and opinion terms) as candidates for the subsequent step. In the second step, these candidates are mapped and ultimately refined into relevant aspect-sentiment triplets. To further enhance the performance of the DGSEP framework, we introduce two innovative strategies, namely DGSEP (S_1) and DGSEP (S_2), which not only significantly improve data augmentation effects but also incorporate different prompt templates. Additionally, we propose a label-oriented sequence label generation fusion module (LSGF), which aims to fuse T5-based and label-oriented sequence labels to improve the ability of the generation model to handle complex structures. Through comprehensive analysis on various benchmarks, we demonstrate that DGSEP achieves state-of-the-art results in nearly all cases.

1. Introduction

In recent years, Aspect-Based Sentiment Analysis (ABSA) has attracted significant attention from researchers as a sophisticated and challenging task in the field of sentiment analysis. The main goal of ABSA is to predict sentiment triplets from a given input text, in order to more comprehensively capture the user's sentiment toward specific aspects. These sentiment triplets typically consist of four key components: aspect terms (a), aspect categories (c), opinion terms (o), and sentiment polarity (s) (Zhang et al., 2022a). Together, these elements form the fundamental units for analyzing sentiment in text and provide insights into the sentiment expressed toward specific aspects.Initially, ABSA research focused on the extraction of individual elements. Specifically, it includes the following directions: Aspect Term Extraction (ATE) (Liu et al., 2015; Ma et al., 2019; Wang et al., 2017), which aims to identify aspects that describe specific entities or attributes in a sentence; Opinion Term Extraction (OTE) (Wang et al., 2017; Wu et al., 2020a; Yu et al., 2018), which focuses on extracting sentiment-related expressions or modifiers in a sentence; and Aspect Sentiment Classification (ASC) (Chen et al., 2020; Lu et al., 2022; Yang et al., 2023b), which determines the sentiment polarity (positive, negative, or neutral) of a given aspect term. These foundational tasks have provided a solid theoretical base for the deeper exploration of ABSA and have promoted its application and development in complex sentiment analysis scenarios.

As research in ABSA continues to evolve, the emphasis has gradually moved beyond the extraction of isolated elements toward tackling more complex tasks, such as triplet and quadruple extraction. This transition reflects a growing need for more fine-grained sentiment analysis and a deeper understanding of nuanced sentiment relationships within textual data. In response, several advanced tasks have been introduced in recent years–such as Aspect Sentiment Triplet Extraction (ASTE) (Peng et al., 2020), Target Aspect Sentiment Detection (TASD) (Wan et al., 2020), Aspect Sentiment Quad Prediction (ASQP) (Zhang et al., 2021a), and Aspect Category Opinion Sentiment (ACOS) (Cai et al., 2020)–each addressing specific subdomains of ABSA and contributing to the expansion of sentiment analysis research.

Among these tasks, Aspect Sentiment Triplet Extraction (ASTE) has gained significant attention due to its research importance and the technical challenges it presents. The goal of ASTE is to simultaneously extract three key components from the text: aspect terms, opinion terms,

* Corresponding author.
E-mail addresses: 2664223216@qq.com (Y. Chen), tang4415@126.com (M. Tang), 1448285452@qq.com (S. Du), whichone276@gmail.com (K. Yang), 2746557318@qq.com (Y. Zheng), zhaomingfeng@cmdi.chinamobile.com (M. Zhao).

https://doi.org/10.1016/j.eswa.2025.130088
Received 30 March 2025; Received in revised form 15 September 2025; Accepted 17 October 2025
Available online 22 October 2025

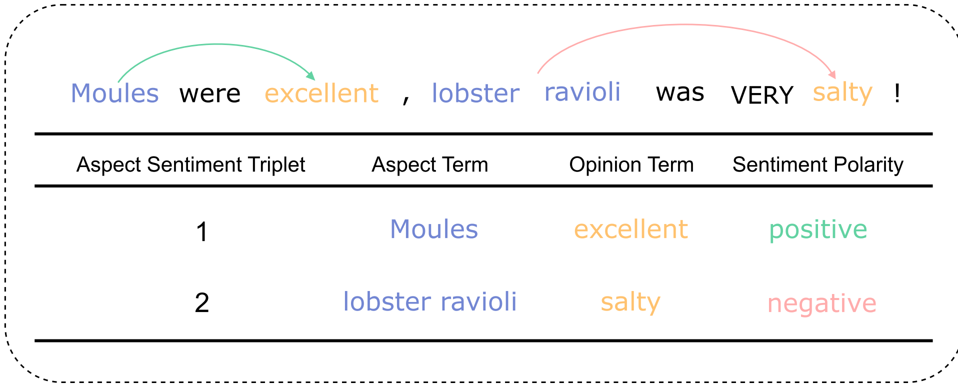

Fig. 1. Example of the ASTE task. Aspect terms and opinion terms are highlighted in yellow and blue, respectively. The green and red arrows from the aspect terms to the opinion terms represent positive and negative sentiments, respectively.

and sentiment polarity. This task requires the system not only to accurately identify the core elements in the text but also to determine the sentiment relationships between these elements. The complexity of ASTE is reflected in two main aspects: (i) the correct extraction of aspect and opinion terms, requiring a deep understanding of the text's semantics; (ii) the accurate identification of sentiment polarity associations between these terms, involving contextual dependence and the diversity of sentiment expression.

As shown in Fig. 1, in the example of the ASTE task, aspect terms and opinion terms in the text are marked in blue and yellow, respectively. For instance, “excellent” is an opinion term associated with the aspect “moules”, expressing a positive sentiment, while “salty” is an opinion term associated with the aspect “lobster ravioli”, expressing a negative sentiment. The sentiment relationship between these terms not only reflects the user's attitude toward specific aspects but also provides valuable insights for product improvement and user demand analysis.

To address the challenges of ASTE, many studies have proposed some solutions, which can be roughly divided into four categories: (1) pipeline methods, (2) end-to-end methods, (3) two-stage methods, and (4) large language model-based methods. The ASTE task initially adopted a pipeline-based approach to extract triplets. In this framework, the first stage identifies the three components of the triplet (aspect, opinion, and sentiment) respectively, and then the second stage uses a binary classifier to pair aspect-opinion candidate words and determine the sentiment polarity (Huang et al., 2021; Peng et al., 2020). Although this method has a clear structure and is easy to implement, it is easy to cause errors to cascade between stages due to the independent modeling of each stage, and it is difficult to capture the complex semantic dependencies between the elements within the triplet, especially when dealing with overlapping triplets. To this end, researchers turned to end-to-end sentiment triplet extraction methods (Wu et al., 2020b; Xu et al., 2025b; Yang et al., 2025b; Zhang et al., 2020; Zhao et al., 2020), which can jointly extract triplets and avoid error accumulation. Among them, the grid annotation scheme GTS (Wu et al., 2020b) performs outstandingly, using a unified grid annotation method to extract all triplets at once. However, GTS still has limitations: it annotates at the word level, splits multi-word items, and only models the relationship between words, resulting in the separation of the semantic information of multi-word aspect words and opinion words. In recent years, two-stage methods and large language model-based methods have been widely introduced in ASTE research. In the two-stage method, independent sequence labeling models are usually used to identify potential aspect and opinion words from sentences respectively, and then the corresponding sentiment polarity is predicted based on the extracted aspects alone (Peng et al., 2020). To address the limitations in generating diverse opinion and sentiment words, Xu et al. proposed a dual-enhanced generative model (GAC) based on graph attention network and contrastive learning (Xu et al., 2024a). To address the limitations in recognizing long-span text and the incomplete identification of complete triplets, Yang et al. proposed a multi-prompt generative model (MPGM) based on self-supervised contrastive learning (Yang et al., 2025a). Since these two stages are independent of each other and do not share information, the overall performance is often limited in actual triplet extraction. To overcome this problem, more and more researchers have begun to adopt large language model-based methods. Methods based on LLM have powerful language modeling capabilities and support flexible end-to-end generation of ASTE tasks. However, they often have difficulty in dealing with structured prediction problems, such as span boundary accuracy, format consistency, and structural constraint satisfaction. These limitations highlight the need to enhance control mechanisms to fully tap the potential of LLM in structured sentiment analysis.

In recent years, an increasing number of studies have explored two-stage methods and the LLM-driven ASTE framework, showing the advantages of guided decoding and multi-view prompts. Zhang et al. (2025) effectively improves the ability to model dependencies between elements in triplets by constructing a graph convolutional network that integrates syntactic and semantic information. However, due to the two-stage structure, it still faces the limitations of error propagation and dependence on external parsing tools. Different from these existing methods, our method adopts a modular two-stage design to strike a balance between performance and interpretability. Specifically, the two-stage generation is guided by individual elements to construct aspect sentiment triplets from multiple semantic paths. Inspired by human intuition in solving complex problems, the task is decomposed into smaller subtasks, which improves flexibility and decoding efficiency. This design also enables the model to better capture the internal dependencies between triplet components and more effectively adapt to the linguistic diversity of natural language.

To address the above challenges and build on our research findings, we propose the DGSEP framework, which adopts a dual-stage process. In the first stage, our method predicts each individual element independently, generating candidates for the subsequent stage. In the second stage, these candidates are mapped and refined to form the final triplets. Within the DGSEP framework, we tailor the modeling process for the ASTE task to achieve two core objectives: (i) accurately predicting the number of individual elements, and (ii) effectively mapping and assembling triplets from the predicted elements. By decomposing ASTE into sequential subtasks, DGSEP enhances task modularity, allowing specialized models to be trained for each step. This design improves prediction accuracy by filtering out irrelevant candidates and progressively refining triplet construction. Additionally, DGSEP increases flexibility in handling complex inputs by generating multiple prediction paths from different starting points and selecting the most probable path through aggregation.

To further promote the implementation of the DGSEP method, we designed two innovative strategies: S_1 and S_2, which not only greatly improve the effect of data augmentation, but also adopt different prompt templates. These prompt templates embed prior knowledge about the role relationship and dependency structure between aspect, opinion and sentiment elements in natural language. By adopting element-first prompting methods (aspect-first, opinion-first or sentiment-first), the generation process can be consistent with the cognitive path of humans when dealing with structured language tasks. The generation of the model is not only based on the input text, but also guided by the semantic roles explicitly defined in the prompts, thereby strengthening the modeling ability of the task structure. This approach can control the generation diversity through different prompt paths, improve the coverage and completeness of triplets, and enhance the interpretability of the results. In addition, the distinction between "→" and "⇒" reflects the ability to model one-to-one and one-to-many relationships in the triplet reasoning process, thereby implicitly introducing logical constraints between elements.

In addition, we propose a label-oriented sequence label generation fusion module (LSGF), which aims to fuse T5-based and label-oriented sequence labels to improve the ability of the generation model to handle complex structures. Specifically, decoding occurs after processing by the LSGF module. The BIO tags generated by the LSGF module help integrate span boundary awareness into the generation framework, so that the boundaries of multi-word aspect/opinion terms can be handled flexibly and efficiently. While the generation model requires multiple calls to the output template of multiple triplet sentences, in LSGF, we use different tag vectors for the same label to avoid confusion and achieve label sharing. Since these labels encode the information of multiple triplets in a sentence, the previously generated labels will help in the decoding of subsequent triplets. In summary, our paper makes the following key contributions:

- We introduce DGSEP, a simple yet effective dual-stage hinting framework, where the first step predicts a single element and the second step maps the single element to another.
- A label-oriented sequence label generation fusion module (LSGF) is proposed, aiming to fuse T5-based and label-oriented sequence labels to improve the capacity of generative models in handling complex structures.
- We propose two different innovative strategies to significantly improve the effect of data augmentation and thus improve the prediction results.
- We conduct extensive experiments on benchmark datasets. The experimental results demonstrate the effectiveness of our model.

2. Related work

As a rapidly emerging subtask in Aspect Based Sentiment Analysis (ABSA), Aspect Sentiment Triplet Extraction (ASTE) has garnered increasing attention from both academia and industry. ASTE aims to provide a comprehensive and structured approach to ABSA by extracting aspect terms, opinion terms, and their associated sentiment polarities. Unlike traditional sentiment analysis tasks, ASTE not only requires identifying key sentiment elements within a sentence but also demands accurately establishing sentiment relationships between these elements, making it a more complex and challenging problem. To tackle this problem, researchers have introduced various innovative methods, which can be broadly categorized into four main approaches: pipeline methods, end-to-end methods, two-stage methods and large language model-based methods. Each of these methods addresses ASTE from different perspectives, contributing to the advancement of the field. These methods vary in terms of extraction accuracy, model efficiency, and generalization capability, forming the core technical framework for ASTE research.

2.1. Pipeline methods

Pipeline methods decompose the Aspect Sentiment Triplet Extraction (ASTE) task into multiple sequential stages, where triplets are predicted by first identifying their components and then verifying their correctness. Mao et al. (2021) reformulated ASTE as two machine reading comprehension (MRC) subtasks: the first MRC model identifies aspect terms in sentences, while the second MRC model predicts the corresponding opinion-sentiment pairs. Similarly, Chen et al. (2021) proposed a multi-round MRC formulation, where aspect terms and opinion terms are extracted separately, then paired, and finally sentiment polarity classification is performed to form complete aspect sentiment triplets. To better model the bidirectional dependencies between aspect terms and opinion terms, Wu et al. (2021) proposed HAST + LOG, a hierarchical pipeline that first identifies opinion terms based on aspect cues and then jointly extracts aspect-opinion pairs using sequence tags. However, due to the inherent sequential nature of pipeline methods, errors in early stages will inevitably propagate to subsequent stages, severely affecting the overall performance. The error accumulation problem remains a fundamental limitation of pipeline-based methods in ASTE. Moreover, pipeline methods still lack the problem of rich element-wise information interactions between triplets.

2.2. End-to-end methods

Compared with pipeline methods, end-to-end methods integrate all subtasks into a unified model, thereby achieving joint context modeling, reducing error propagation between modules, and improving the extraction of implicit sentiment and complex semantic dependencies within sentences. Li et al. (2022) proposed a span-based framework that enumerates all possible spans between words to identify aspect terms and opinion terms. However, this approach may lead to span overlap or redundant conflicts. To address this issue, Gao et al. (2023) proposed a multi-level joint triplet extraction strategy, which first identifies all potential relation types through multi-label classification, and then performs span selection and relation graph construction under each relation type, thereby effectively alleviating the problem of triplet overlap. In span-based methods, exhaustive span enumeration may generate a large number of invalid triplet candidates. To address this issue, Yang et al. (2023c) designed a bidirectional triplet extraction mechanism that simultaneously considers the forward and backward interactions between spans, combines domain knowledge to improve semantic representation, and applies intersection-based filtering to remove incorrect triplets. Joint extraction methods also belong to the category of end-to-end frameworks. For example, Wu et al. (2020b) proposed a Grid Tagging Scheme (GTS), which uses a unified tagging structure to extract all triplet elements simultaneously, thereby improving the extraction accuracy and achieving good results. Chen et al. (2022) proposed an enhanced multi-channel graph convolutional network model, in which adjacent words and relation tensors are regarded as nodes and edges to construct a multi-channel graph, so that the model can fuse various language features in a structured manner. In addition, Zhang et al. (2022b) proposed a Boundary Driven Table Filling (BDTF) method, which represents each triplet as a relation region in a two-dimensional table, effectively converting the ASTE task into a region detection and classification problem.

2.3. Two-stage methods

The two-stage methods decompose the ASTE task into two sequential subtasks. The first stage focuses on information extraction, typically identifying aspect terms and opinion terms from text. The second stage classifies the sentiment polarity (positive, negative, or neutral) of the extracted aspect-opinion pairs. By decoupling extraction and classification, this approach simplifies model design and allows specialized architectures to be used for each subtask. However, by training independently, the two stages may lead to error propagation, where errors in the first stage may adversely affect the performance of the second stage. Peng et al. (2020) first formally defined the ASTE task and proposed a two-stage pipeline to solve it. In their approach, two independent sequence labeling models were used in the first stage to extract aspect terms and corresponding opinion terms containing sentiment information. Then, in the second stage, these candidate models were paired and a classifier was used to determine the validity of each triplet. Li et al. (2023b) further improved on this by proposing a dual-strategy network (DPM-MTR) consisting of an upper annotation network for extracting aspect terms and a lower network for extracting opinion terms and performing sentiment classification based on aspect-specific representations. In related work, they also designed a three-stage sequential annotation strategy that sequentially extracts aspect terms and opinion terms, followed by sentiment classification, also based on aspect-aware representations. Despite the progress, the two-stage approach still faces several limitations, including inherent risk of cascading errors, increased architectural complexity, and insufficient modeling of the complex dependencies between aspect terms, opinion terms, and sentiment during the extraction process (Xu et al., 2021).

2.4. Large language model-based methods

With the rapid development of large language models (LLMs), recent studies have begun to explore their potential in solving the more challenging Aspect Sentiment Triplet Extraction (ASTE) task. ASTE requires the simultaneous identification of aspect terms, opinion terms, and corresponding sentiment polarity. Zhang et al. (2021c) formulated ASTE as a structured generation problem and used a T5-based model to directly generate complete sentiment triplets from raw text, demonstrating the feasibility of applying sequence-to-sequence LLM to structured sentiment extraction. Lu et al. (2025) proposed the QAIE framework, which introduced quantity-aware and information-enhanced modules to enhance the generative model's understanding of ABSA rules, thereby improving performance in few-shot scenarios. Zhang et al. (2024) conducted an empirical evaluation of the performance of LLM in sentiment analysis and pointed out that although the general LLM can complete complex tasks such as ABSA under zero fine-tuning conditions, its performance is still inferior to that of specially trained models. Simmering and Huoviala (2023) compared the performance differences between GPT-4 and GPT-3.5 in the ABSA task under zero-shot, few-shot, and fine-tuning settings. To improve domain adaptability, Xu et al. (2024b) proposed to fuse BERT with LLM to leverage the strengths of both to enhance the performance of Chinese ABSA. These LLM-based methods benefit from powerful language modeling capabilities and flexible generation capabilities, but most existing studies do not explicitly address the performance degradation caused by sparse input information or data scarcity (which often occurs in the real world). These limitations highlight the need for more structure-aware and robust LLM-based ASTE methods, especially in resource-poor or noisy environments.

3. Method

The following subsections explain the details of DGSEP and the two different innovation strategies. An overview of DGSEP is shown in Fig. 2.

3.1. Problem statement

For an input sentence $H = \{(k_1, k_2, \cdots, k_n)\}$, which consists of n words ($n \geq 1$), the ASTE task aims to directly generate all the triplets contained in the given sentence H. The number of triplets in the sentence is m, and the i triplet can be expressed as $m_i = (a_i, o_i, s_i)$, where a_i and o_i represent the aspect item and opinion item in the sentence H, respectively, and s_i is the sentiment polarity label with a value range of $\{POS, NEU, NEG\}$, corresponding to positive, neutral, and negative sentiments, respectively. Therefore, the goal of the model is to identify all the triplets contained in the sentence H, and provide an aspect term, an opinion term, and its corresponding sentiment polarity for each triplet.

3.2. Training

3.2.1. First step input and target

The goal of this step is to predict the individual elements such as aspect terms, opinion terms, and sentiment polarity, respectively, as shown in Fig. 3. Assuming that m represents the total number of different element types in the ASTE task, a new set of data points $(r_j, u_j)_{j=1}^{m}$ can be generated from the input sentence X. At the same time, Q and A are defined as $Q = \{g_1 : \}\}aspect'', g_2 : \}\}opinion'', g_3 : \}\}emotion''\}$ and $A = \{g_1 : \}\}iLIfe, SnowLeopardX'', g_2 : \}\}great, love'', g_3 : \}\}positive, positive''\}$ respectively, where the input and target structures corresponding to each j from 1 to m are given in Eq. (1).

$$\begin{aligned} x_i &: x \Rightarrow what\ Q[t_i]? \\ y_i &: A[t_i] \end{aligned} \tag{1}$$

After annotating each data point in the original dataset, a final dataset D_1 is generated. The model is then fine-tuned on D_1 for the seq2seq task, further advancing the training of the initial model.

3.2.2. Label-oriented sequence label generation fusion module(LSGF)

Although tag-based templates can guide the generation model by incorporating aspect, opinion, and sentiment tags, the traditional encoder-decoder architecture has significant limitations in generating complex structured outputs. This is specifically manifested in two aspects: first, the token-by-token generation mechanism cannot effectively define the boundary information of multi-word semantic units (such as compound aspect/opinion terms); second, when processing sentences containing multiple triplets, the strategy of repeatedly calling fixed templates is prone to semantic cross-interference, thereby reducing the accuracy of generated content. To address the above issues, this study designed a label-oriented sequence label generation fusion module (LSGF), which embeds structured prompt information such as aspect terms, opinion expressions, and sentiment polarity into the generation process to achieve end-to-end sequence annotation output. As shown in Fig. 4, this module significantly improves the generation quality of complex structured text by dynamically integrating semantic tags and context features.

When faced with multi-word aspect or opinion terms, traditional span-based methods directly enumerate candidate spans, but are limited by the maximum span length and inefficient due to a large number of invalid spans. In contrast, our proposed label-oriented sequence label generation fusion module (LSGF) processes multi-word expressions at the token level while preserving span awareness through BIO annotations. Specifically, LSGF generates token-level BIO tags for aspect and opinion terms before generation, allowing the model to pre-identify span boundaries without exhaustive enumeration. These span boundaries are encoded into the input of the generative model, guiding it to generate structurally consistent triplets. Compared with span-based models, LSGF does not rely on fixed-length span candidates, thus avoiding the strict restriction of span length. It can flexibly capture longer or nested aspect/opinion phrases without increasing computational overhead. In addition, the label vectors used in LSGF maintain structural consistency between decoding steps, enabling label sharing and boundary enforcement, which is particularly useful in sentences containing multiple or overlapping multi-word terms.

Before performing the LSGF module, we first input the sentence X into the Transformer encoder to obtain the context feature K^{enc}, as detailed in Eq. (2).

$$\mathbf{K}^{\mathbf{enc}} = \mathbf{Encoder}(X_a) \tag{2}$$

Then in LSGF, we adopt a sequence labeling approach to identify aspect and opinion terms. Specifically, we first extract

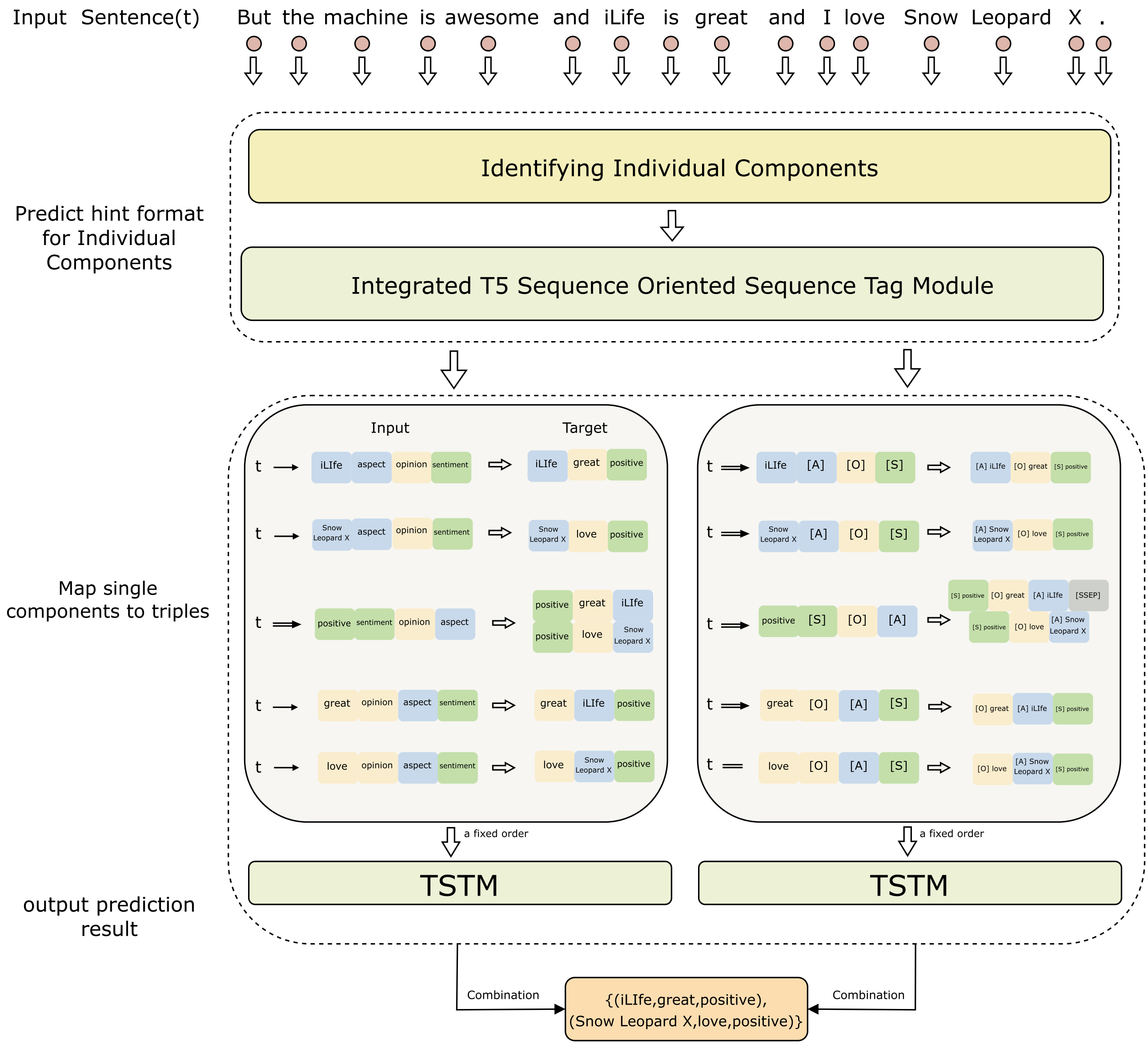

Fig. 2. The overall preview of Dual-stage generative model with sequence-oriented labeling and element-to-tuple prompting improves aspect sentiment triplet extraction(DGSEP).

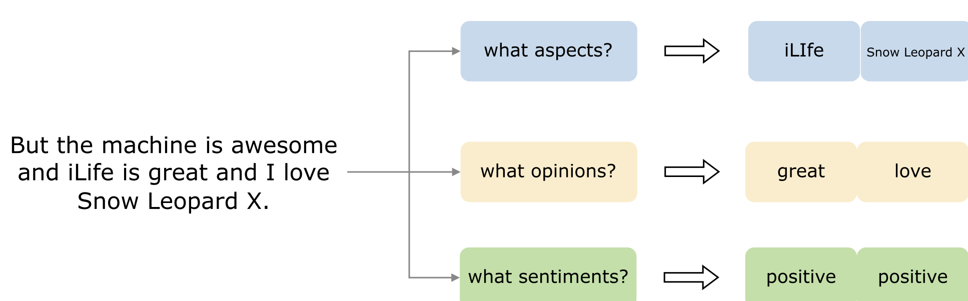

Fig. 3. An overview of the first step input and goals.

aspect features $\mathbf{K}^a = \{\mathbf{k}_1^a, \mathbf{k}_2^a, \cdots, \mathbf{k}_L^a\} \in \mathbb{R}^{U\times v}$ and opinion features $\mathbf{K}^o = \{\mathbf{k}_1^o, \mathbf{k}_2^o, \cdots, \mathbf{k}_U^o\} \in \mathbb{R}^{U\times v}$ from the context features K^{enc} through two linear transformations, where U represents the length of the sentence, as detailed in Eq. (3).

$$\mathbf{K}^a = \mathrm{MLP}_a(\mathbf{K}^{\mathbf{enc}}), \mathbf{K}^o = \mathrm{MLP}_o(\mathbf{K}^{\mathbf{enc}}) \tag{3}$$

We use a transform decoder to generate the target sequence Y. At time step t, the decoder calculates the current hidden state k_t based on the context feature K^{enc} and the previously decoded token sequence $y_{[1:t-1]}$, as detailed in Eq. (4).

$$\mathbf{k}_t = \mathbf{Decoder}(y_{[1:t-1]}, \mathbf{K}^{\mathbf{enc}}) \tag{4}$$

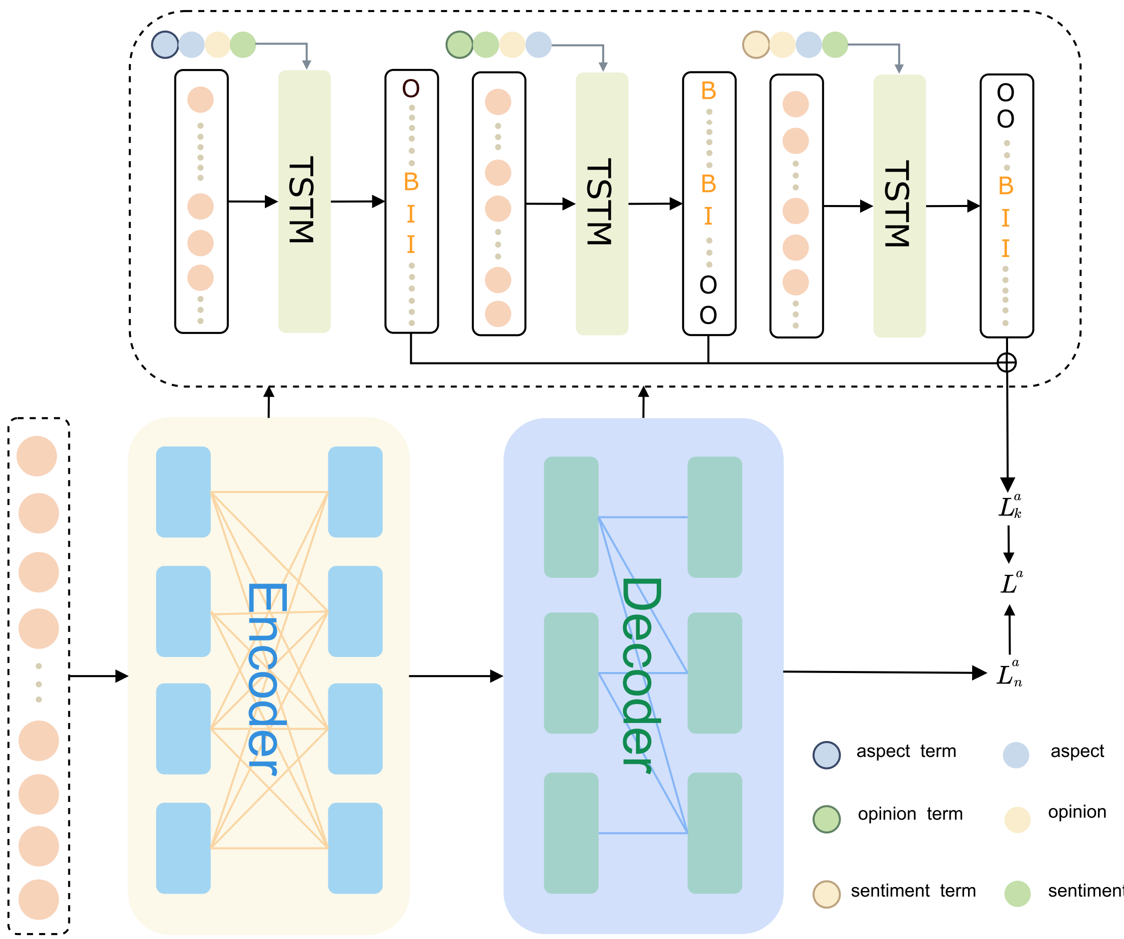

Fig. 4. The second step is to enhance the process of converting a single element to a tuple through the LSGF module.

Next, k_t is used to calculate the conditional probability of labeling y_t, as detailed in Eq. (5).

$$c(y_t|\mathbf{K}^{\mathbf{enc}}; y_{[1:t-1]}) = softmax(\mathbf{W}^{\mathbf{T}}\mathbf{k}_t) \tag{5}$$

where W is the transformation matrix. Finally, we calculate the cross entropy loss $\mathcal{L}_g^{a\to o}$ between the decoder output and the target sequence Y, as detailed in Eq. (6).

$$\mathcal{L}_g^{a\to o} = -\sum_{i=1}^{U} \log c(y_t|\mathbf{K}^{\mathbf{enc}}; y_{[1:t-1]}) \tag{6}$$

Subsequently, we use the last hidden state of the decoder as the basis for the tag features, and obtain the aspect tag features $\mathcal{P}^a = \{\mathrm{p}_1^a, \mathrm{p}_2^a, \cdots, \mathrm{p}_J^a\}$ (where J represents the number of triplets) and the opinion tag features $\mathcal{P}^o = \{\mathrm{p}_1^o, \mathrm{p}_2^o, \cdots, \mathrm{p}_J^o\}$. Next, we perform sequence labeling by calculating the tag-oriented features (e.g.,p_i^a or p_i^o), as detailed in Eq. (7).

$$\begin{aligned} \mathbf{z}_{ij}^a &= \sigma(\mathbf{W}_1(\mathbf{k}_j^a \oplus \mathbf{p}_i^a) + \mathbf{b}_1) \\ \mathbf{z}_{ij}^o &= \sigma(\mathbf{W}_1(\mathbf{k}_j^o \oplus \mathbf{p}_i^o) + \mathbf{b}_1) \end{aligned} \tag{7}$$

where $\sigma(\cdot)$ is the selu activation function, $\mathbf{k}_j^a \in \mathbf{K}^a$ and $\mathbf{k}_j^o \in \mathbf{K}^o$ are aspect and opinion features. W and b are the transformation matrix and bias.

In the training phase, we adopt a "label first, generate later" strategy. Specifically, the LSGF module first predicts BIO labels for the to-be-annotated words in a sentence, and then the generative model generates text based on these labels. This approach helps guide the text generation module to better discern the boundary information of multi-word opinion or sentiment expressions. The boundary enforcement mechanism proposed in this paper ensures consistent segmentation of multi-word aspect and opinion terms during the decoding step, significantly reducing span prediction ambiguity and avoiding duplicate or inconsistent annotations.

When processing an input sentence containing multiple triplets, the LSGF module generates a different tag sequence for each aspect/opinion tag feature. For example, $G_i^{pa} = \{y_{i1}^{pa}, y_{i2}^{pa}, \cdots, y_{iL}^{pa}\}$ represents the aspect term p_i^a, and $G_i^{po} = \{y_{i1}^{po}, y_{i2}^{po}, \cdots, y_{iL}^{po}\}$ represents the opinion term p_i^o, where G^{pa} and G^{po} are BIO tags in the sequence tagging. In the generation module, the same tag can share information without causing confusion because it points to different pointers to multiple aspect/opinion terms in LSGF, which helps to decode sentences containing multiple triplets. Next, we input the tag-oriented features into the fully connected layer to predict the tags of the aspect/opinion terms and obtain the predicted probabilities for the tag set, as detailed in Eq. (8).

The interpretability of LSGF lies in the principles behind its label sharing and boundary enforcement mechanisms. Label sharing enables the decoding process to reuse previously predicted boundary representations, which provides a transparent explanation of how to maintain a consistent multi-word span across different decoding steps. In turn, boundary execution guarantees that span boundaries follow BIO constraints rather than arbitrary splits, thus ensuring that split and merge operations are performed in a rule-based and interpretable manner rather than opaque decisions of the model.

$$\begin{aligned} f_{ij}^{ma} &= softmax(\mathbf{W}_2\mathbf{z}_{ij}^a + \mathbf{b}_2) \\ f_{ij}^{mo} &= softmax(\mathbf{W}_2\mathbf{z}_{ij}^o + \mathbf{b}_2) \end{aligned} \tag{8}$$

The training loss of LSGF is defined as the cross entropy loss, as detailed in Eq. (9).

$$\mathcal{L}_m^{a\to o} = -\sum_{i=1}^{N}\sum_{j=1}^{U}\sum_{c\in E}\mathbb{I}(y_{ij}^{ma}=c)\cdot log(f_{i,j|c}^{ma}) - \sum_{i=1}^{N}\sum_{j=1}^{U}\sum_{c\in E}\mathbb{I}(y_{ij}^{mo}=c)\cdot log(f_{i,j|c}^{mo}), \tag{9}$$

where $I(\cdot)$ is the indicator function, y_{ij}^{ma} and y_{ij}^{ma} are the ground truth labels, and E represents the set of {B, I, O} labels.

3.2.3. Second step input and target

In the second phase, we design two innovative strategies. Similar to the first phase, the training dataset consists of pairs of $\{(x_i, y_i)\}_{i=1}^{N}$, where N represents the total number of data points, and x_i and y_i represent the input text and the target text, respectively. It is important to note that this step is independent of the first phase, allowing the two models to be trained simultaneously, thus avoiding the increase in time complexity. In this phase, for each input labeled x_i, we assign y_i to a set $T = \{(a_i, o_i, s_i)\}_{i=1}^{|T|}$. For each x_i, we construct $L = \{a_i\}_{i=1}^{|T|} \cup \{o_i\}_{i=1}^{|T|} \cup \{s_i\}_{i=1}^{|T|}$, forming a set of unique elements. During inference, individual elements are accessed instead of tuples, so that the set L can be directly formed through the predictions of the first phase. First, we outline the prompt template we designed and introduce a fixed order for the prompt elements to be arranged, and then compare the two innovative strategies.

First Template (T_1): In the triplet prediction task, the task prompt (t_p) is defined as a sequence of different element types corresponding to each element type $l \in L$. The set V is initialized by the existing element labels in the task:$V = \{aspect, opinion, sentiment\}$. If the element type t_l is not the first item, its position will be swapped with the first item to ensure the priority order. Therefore, the task prompt tp is also defined accordingly. Next, the input x_e of the new dataset is determined, as detailed in Eq. (10).

$$x_e := x_i \to l : t_p \tag{10}$$

To ensure that the element l in the task prompt (t_p) is always the initial element of all prediction tuples, we put l at the front of the task prompt to ensure its first position in the task prompt. In addition, when l appears multiple times in the set T, we replace the symbol "→" with "⇒" to distinguish the different meanings of the symbol during reasoning. Specifically, when encountering "→", the model generates a single tuple; when encountering "⇒", the model generates multiple tuples to reflect the indication of multiplicity.

Second Template (T_2): Compared with the previous template, this template has two main differences. First, it only uses the "⇒" symbol instead of "→". Second, it adopts the marking method proposed by Gou et al. (2023) to indicate the element type: [A] represents aspect words, [O] represents opinion words, and [S] represents sentiment polarity. Each output element is preceded by a corresponding marker, and different tuples are separated by "[SSEP]".

A Fixed Order: After constructing a new dataset named D' using the first or second template, the second step of the DGSEP input reconstruction method consists of anchoring the first prompt element in the task prompt and then arranging the remaining elements to generate new commands in the same task prompt.

In the ASTE task, when the task prompt starts with "aspect" or with a fixed marker "[A]", two enhanced modes can be generated based on the prompt design scheme of template T_1 or T_2: "aspect, opinion, sentiment" and "aspect, sentiment, opinion" or "[A][O][S]" and "[A][S][O]". By appending these sequences after the input text, each data point can be double-labeled. It is worth noting that although the reasoning stage can only adjust the order of elements of the input component, the arrangement of the target elements will keep pace with the task prompt during training. For the aspect sentiment triplet extraction task, there are (3-1)! = 2 possible permutations. This study uses a systematic permutation and combination strategy to construct the D_2 dataset as the basis for the second stage of training, and finally completes the fine-tuning of the T5 model on this dataset. The implementation path of this method can be summarized as: based on permutations and combinations, data is selected from potential possibilities to construct D_2, and then model optimization is carried out. This method is as follows:

All Selection: This method achieves a higher data augmentation rate by considering all possible permutations and combinations of each data point in the set D'.

3.2.4. Innovative strategies

This section introduces two innovative strategies applied in the study, using task hints and selection methods for data augmentation to achieve the core goal. The goal of these strategies is to improve the effectiveness of the tuple prediction model.

- **The first innovation strategy** (S_1): In this innovative strategy, the all-selection method is adopted under the T_1 template, named S_1.
- **The second innovation strategy** (S_2): This innovative strategy utilizes the T_2 template and performs a select-all approach, named S_2.

3.2.5. Loss function

DGSEP adopts a dual-stage process to accurately predict sentiment tuples. It involves designing different objectives and inputs for each step. Each step uses the LSGF model, which is fine-tuned independently as the main model to minimize the cross entropy loss function during training, as detailed in Eq. (11).

$$\mathcal{L} = \lambda(\mathcal{L}_g^{a\to o} + \mathcal{L}_m^{a\to o}) + (1-\lambda)(\mathcal{L}_g^{o\to a} + \mathcal{L}_m^{o\to a}) \tag{11}$$

Here, λ is a hyperparameter that controls the contribution of different strategies.

3.2.6. Constraint decoding(CD)

During inference, we adopt a constrained decoding (CD) strategy to ensure that the generated content and format meet the legality requirements. This strategy is inspired by Bao et al. (2022) and Lu et al. (2021). Content legality means that the aspect or opinion terms should be single or multiple consecutive words in the input sentence, and the sentiment should be positive, neutral, or negative; format legality means that the generated sequence should meet the format requirements defined by the template.

During the decoding process, we consider two types of legitimacy as constraints on the candidate vocabulary. Before starting decoding, we list the candidate vocabulary for each token in the input sentence and template. Next, based on the current input token, the candidate vocabulary is adjusted in each decoding step. For example, when the "</s>" start token is input to the decoder, the candidate token should be "aspect" or "opinion" to ensure format legitimacy; when ":" is input, the model needs to determine the first word of the aspect or opinion term, and the candidate token should be consistent with the word in the input sentence.

3.2.7. Inference

The inference process is different from the training process, in which the two models are run in conjunction with each other. The initial model first predicts a single element and uses lightweight constrained decoding to ensure that the correct token type is generated. Based on this, a new dataset D_1' is generated for further evaluation in the second step.

To continue the evaluation, we choose D_1' as the dataset and construct the input of the second step based on different methods of the second step. Then, the output of the second step model is used to generate D_2'. According to the input design in the previous part, each input in D_2' consists of multiple parts: the review sentence (denoted by X), the task prompt (denoted by l), and the symbol "→" or "⇒". After examining D_2', we grouped the dataset according to the shared sentences and designed an aggregation method for each group, which can be generalized to finally determine the output of other groups. Within each group, we verify the generated elements according to the syntactic format of

Table 1
Statistics of the ASTE-Data-V2 dataset are shown here. # indicates the number of items. Sent, Asp, Opn, and Tri indicate sentences, aspects, opinions, and triplets, respectively. #Tri overlap indicates sentences containing overlapping triplets. #TM indicates the number of triplets where at least one aspect/opinion term contains multiple words.

Dataset		#Sent	#Asp	#Opn	#Tri	#Tri overlap	#TM
14lap	Train	906	1280	1254	1460	257	636
	Dev	219	295	302	346	59	156
	Test	328	463	466	543	97	252
14res	Train	1266	2051	2061	2338	367	752
	Dev	310	500	497	577	101	189
	Test	492	848	844	994	174	337
15res	Train	605	862	935	1013	154	335
	Dev	148	213	236	249	44	84
	Test	322	432	460	485	66	188
16res	Train	857	1198	1300	1394	210	476
	Dev	210	296	319	339	52	123
	Test	326	452	474	514	76	170

the specified prompt template and eliminate the parts that do not meet the requirements. Then, we denote the input within the group as w, and initialize the set T'_w for each input, and assign the generated tuples to the corresponding T'_w. Finally, the aggregation is completed through Eq. (12), where T' represents the final output of a specific review sentence in each group.

$$T' = \left\{ t \in \bigcup_{i=1}^{k} T'_i \mid \sum_{i=1}^{k} l_{T'_i}(t) > d \right\} \tag{12}$$

The formula contains a variable d that can be adjusted according to the task requirements. Although the hyperparameter D is not explicitly defined in the formula, it is used to modify the entire process. If T' is still empty after applying the formula, the value of d is reduced by one and the process is repeated. If T' is still empty, it will proceed to the maximum number of iterations of D.

4. Experiment

4.1. Datasets

We evaluated the model on ASTE-DATA-V2, which contains four datasets. In the following sections, we refer to them as 14res, 15res, 16res, and 14lap. The datasets in the restaurant domain are 14res, 15res, and 16res, while 14lap is in the laptop domain. In these datasets, each review is annotated with aspect terms, opinion terms, and sentiment polarity. They were created by Peng et al. (2020) based on SemEval Challenges. Later, in order to improve the quality of the datasets, Xu et al. (2020) made significant improvements in the quality and completeness of the annotations, thereby more comprehensively covering the actual situation and enhancing the utility and completeness of the datasets. More details about these datasets are shown in Table 1.

4.2. Baseline

To verify the effectiveness of the proposed method, we conduct a comprehensive comparison with leading baseline methods previously designed for ASTE. We classify the benchmarks into four categories: pipeline methods, end-to-end methods, two-stage methods, and large language model-based methods. The following is a brief description of some of the benchmark methods we selected:

1. **Pipeline Methods**
   - **DE-OTE-BISDD** (Dai et al., 2022) proposes a framework that combines two-way vector embedding with a bidirectional sentiment correlation analysis module. By introducing a deep bidirectional sentiment dependency detection mechanism, it simultaneously optimizes the implicit correlation modeling between text representation and sentiment elements.
   - **HIM** (Liu et al., 2023) designs a hierarchical two-way interactive architecture and used the multi-task collaborative training framework (MTL) to dynamically fuse task-specific features with cross-task common representations to enhance the correlation modeling of multiple subtasks in Aspect Sentiment Triplet Extraction (ASTE).
   - **Dual-MRC** (Mao et al., 2021) proposes a collaborative architecture based on dual-path machine reading comprehension (MRC). Through the joint optimization of the BERT encoding layer, LSTM sequence modeling and self-attention mechanism, it decouples features for sentiment polarity discrimination and entity association reasoning, respectively, to achieve a collaborative enhancement of end-to-end information extraction performance.
2. **End-to-end Methods**
   - **EMC-GCN** (Chen et al., 2022) is a deep learning framework designed for the Aspect Sentiment Triplet Extraction (ASTE) task. Based on a multi-channel GCN structure, the model simultaneously models the syntactic associations and semantic dependencies between words, integrates the grammatical features of the text and the deep semantic representation, and achieves efficient joint extraction of aspect entities, corresponding opinion expressions, and their sentiment polarity triplets. Its core mechanism uses a multi-dimensional relational encoding strategy to effectively improve the model's ability to parse fine-grained sentiment elements and their interactive relationships in complex language structures.
   - **DGEIAN** (Shi et al., 2022) proposes an architecture based on a bidirectional interactive attention mechanism. By integrating a bidirectional long short-term memory network (Bi-LSTM) and a stacked graph convolutional network (GCN), it collaboratively models the syntactic dependency tree structure features of text and the dynamic representation of context sequences, thus achieving joint reasoning of syntactic constraints and semantic clues. This design effectively bridges the advantages of sequence modeling and graph structure analysis through a hierarchical cross-modal interaction mechanism.
   - **SSJE** (Li et al., 2022) constructs a syntactic-semantic joint representation enhancement framework, which modeled the syntactic dependency tree structure features through a graph convolutional network (GCN), combined with a self-attention mechanism to capture the dynamic association of contextual semantics, and improved the joint accuracy of sentiment factor parsing and structured extraction.
   - **ESGAT** (Yang et al., 2023a) uses the perturbation masking influence matrix (PM) to generate edge-enhanced sentiment representation, and combines it with the graph attention network (GAT) to dynamically weighted aggregate sentiment-related edge features to mine the complex topological relationships of fine-grained sentiment interactions.
   - **DRN** (Xia et al., 2024) is a method based on sequence labeling. It enhances entity extraction and entity matching tasks respectively through dual relation encoding networks (EER and EMR). EER uses multi-channel graph convolutional networks (GCN) to incorporate semantic and syntactic information, and EMR combines criss-cross attention to capture global interaction information. Finally, entities are extracted through sequence labeling, and aspect sentiment triplets are generated with the help of span contraction labels.
   - **Biston** (Hao et al., 2024) adopts a fusion architecture of a dual-channel heterogeneous encoder (syntactic/semantic dual channels) and a multi-channel graph neural network . Through a heterogeneous graph structure, it simultaneously captured the grammatical dependency features and deep semantic associations of the text, and achieved accurate joint extraction of aspect sentiment triplets.

- **SBRS** (Yang et al., 2024) is a span-based method that first extracts aspect-opinion pairs through a dual-path mechanism, then perceives word pair interaction information in a bidirectional recursive and parallel computing manner, and makes full use of context, semantics, and relational features to classify sentiment polarity, thereby improving the accuracy of the Aspect Sentiment Triplet Extraction (ASTE) task.
- **SATPC** (Li et al., 2024) proposes a span extraction framework based on a part-of-speech filtering module and representation decoupled contrastive learning. It strengthens the recognition of aspect-opinion term boundaries through the part-of-speech filtering mechanism, and uses contrastive learning to distinguish syntactic structure from semantic representation, thereby enhancing the robustness of triplet extraction and its cross-scenario generalization ability.

3. **Two-stage Methods**
   - **SSFF** (Xu et al., 2025a) is a span-based method designed for the Aspect Sentiment Triplet Extraction (ASTE) task. It improves the model's ability to parse aspect terms, opinion terms, and their sentiment polarity by integrating syntactic information (including part-of-speech information for assisting span category prediction and dependency distance information for optimizing sentiment polarity classification). The model improves its adaptability to complex language structures while optimizing the two-stage learning objectives of the span method (i.e., span classification and sentiment polarity classification).
   - **Dual-Span** (Li et al., 2023a) proposes a bidirectional span generation framework, which uses span internal hierarchical association modeling and span-to-span relational reasoning graph network (RGNN), combined with syntactic dependency constraints and part-of-speech correlation features, to dynamically generate candidate span sets to capture the complex interaction patterns between language units.
   - **SSGCN** (Zhang et al., 2025) proposes a GCN method for joint syntactic-semantic graph construction for ASTE tasks. In the first stage, the model extracts aspect and opinion candidates from dependency and semantic relations; in the second stage, polarity classification is performed based on the contextual representation of candidate pairs. Through dual graph encoding, this method effectively integrates language structure and semantic information, improving the extraction performance of complex sentences and overlapping triplets.
   - **IERET** (Li et al., 2025) proposes a two-stage ASTE method that combines implicit expression recognition and table filling. In the first stage, the model recognizes and annotates the emotional expressions that do not appear explicitly but are implicit in the sentence as structural signals; in the second stage, this enhanced information is added to the table filling module to extract triplets by predicting the token-token relationship table.
4. **Large Language Model-based Methods**
   - **BART-ASTE** (Yan et al., 2021) proposes a generative unified paradigm based on BART, reconstructing the ABSA multi-subtask into a sequence-to-sequence end-to-end generation problem, and using the strong semantic generalization ability of the pre-trained generative model to achieve cross-task performance breakthroughs and architectural simplification.
   - **GAS** (Zhang et al., 2021c) proposes a generative unified architecture based on the T5 pre-trained model, which defines the multi-task learning process through the annotation-extraction dual paradigm training strategy and adapts to the joint modeling of ABSA subtasks at different abstraction levels.
   - **PARAPHRASE** (Zhang et al., 2021b) builds a generative framework based on semantic guidance, which generates structured interpretation text by parsing the implicit semantics of natural language tags, and achieves semantic compatibility transfer and adaptation across ABSA tasks while solving ASQP tasks.
   - **EDWPN** (Fei et al., 2021) designs an opinion-aware encoder-decoder pointer network, which explicitly models the multi-granularity interaction pattern of aspect words and opinion words in overlapping structures through a high-order semantic aggregation module, thereby enhancing the boundary perception ability of nested and crossed triplets.
   - **SA-Transformer** (Yuan et al., 2023) focuses on encoding syntactic dependency structures into the Transformer encoder, and proposes a structure-aware mechanism to guide LLM to more accurately model the complex dependencies between aspect terms, opinion terms, and sentiment polarity. This method effectively improves the model's performance in processing overlapping triplets, long-distance dependencies, and low-information sentences without destroying the original LLM architecture, providing a new idea for the integration of structural information and large language modeling capabilities.
   - **ContrASTE** (Sun et al., 2024) proposes an ASTE method that combines minimal label design and contrastive learning, aiming to improve the model's ability to recognize triplet structures under few-sample conditions. Inspired by the few-shot generalization capabilities of large language models (such as GPT), this method effectively enhances the model's modeling of boundaries and syntactic-semantic relationships through structure-guided annotation and contrastive loss of representation decoupling, demonstrating excellent structural perception and cross-scenario generalization capabilities.
   - **QAIE** (Lu et al., 2025) aims to improve the ability of large language models to perform ABSA tasks in few-shot situations. The method consists of two modules: the quantity enhancement module generates additional training samples through prompt guidance to alleviate the problem of sample scarcity; the information enhancement module introduces domain keywords and structural priors to guide the model to focus on key emotional components.
   - **Zero-shot** uses a large language model (LLM) to perform aspect-based sentiment analysis. The specific method is to input a prompt sentence and directly output the corresponding [A, O, S] triplets. The following is an example of text provided to the large language model (LLM) (with the prompt added): "Perform aspect-based sentiment analysis on the provided text and return triplets as [Aspect, Opinion, Sentiment]. You only need to provide the triplets, no additional explanations are required. The provided text: {sentence}"
   - **Few-shots** is based on the zero-shot method, where we add a small number of examples from the training set to the prompt sentence: "Perform aspect-based sentiment analysis on the provided text and return triplets as [Aspect, Opinion, Sentiment]. For example: input: {train sentence} output: {train triplets}, ... (some other examples). You only need to provide the triplets, no additional explanations are required. The provided text: {sentence}"

4.3. Implementation details

To ensure reproducibility and fairness, we use the Hugging Face Transformers library and PyTorch to implement all models. The T5-BASE backbone network (12-layer encoder-decoder, 768 hidden layer size, 12 attention heads, about 220 million parameters) is used in both stages of DGSEP. We use the AdamW optimizer (Loshchilov & Hutter, 2017) with an initial learning rate of 3e-4 and a linear warmup. All models are trained on a single NVIDIA RTX 3090 GPU with 24GB of video memory. The batch size is fixed to 16, and the greedy decoding algorithm is used during inference to ensure the consistency of the evaluation results. The model is trained on the ASTE dataset for 15 epochs. For the inference of the two innovative strategies (S_1 and S_2) we proposed, we set the hyperparameters to d = 3 and D = 1, respectively. The maximum input sequence length is set to 128 and dropout is set to 0.1. The

Table 2
Main results on the ASTE-Data-V2 dataset, the best results are marked in bold and the second best results are underlined. All baseline results are from the original paper.

Model	14lap			14res			15res			16res		
	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)	P(%)	R(%)	F1(%)
Pipeline												
De-Ote-BISDD	56.17	46.20	50.70	68.57	59.17	63.53	61.54	48.43	54.21	65.20	61.34	63.21
HIM	65.99	56.05	60.59	76.99	70.46	73.57	69.65	63.23	66.26	73.11	71.05	72.06
Dual-MRC	57.39	53.88	55.58	71.55	69.14	70.32	63.78	51.87	57.21	68.60	66.24	67.40
End-to-end												
EMC-GCN	61.70	56.26	58.81	71.21	72.39	71.78	61.54	62.47	61.93	65.62	71.30	68.33
DGEIAN	60.15	43.44	51.14	71.68	61.62	66.26	61.84	50.99	55.89	69.40	60.15	64.37
SSJE	67.43	54.71	60.41	73.12	71.43	72.26	63.94	66.17	65.05	70.82	72.00	71.38
ESGAT	61.11	58.96	60.01	72.43	74.05	73.23	65.53	63.50	64.50	67.36	68.81	68.08
DRN	66.99	52.61	58.94	75.24	64.49	69.45	68.61	55.47	61.34	73.30	64.42	68.57
Biston	65.99	53.59	59.15	70.03	68.41	69.21	65.25	56.91	60.79	66.85	69.84	68.32
SBRS	62.30	57.47	59.79	72.08	73.09	72.68	62.50	63.69	63.09	66.92	72.00	69.37
SATPC	65.20	60.18	62.59	73.33	**76.31**	74.79	68.39	61.97	65.03	71.02	71.31	71.17
Two-stage												
SSFF	68.72	59.84	63.97	75.57	75.73	**75.65**	69.23	**69.54**	68.38	74.85	73.04	73.93
Dual-Span	67.14	62.13	64.49	77.01	74.00	75.47	67.97	66.34	67.13	73.56	73.48	73.49
IERET	69.82	56.15	62.24	**78.51**	70.93	74.52	70.27	62.81	66.33	72.26	73.68	72.96
SSGCN	68.56	52.43	59.74	74.63	71.42	73.05	58.28	65.78	61.74	67.59	69.55	68.57
Large Language Model-based												
BART-ASTE	65.52	**64.99**	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
GAS	61.65	58.19	59.87	71.08	71.67	71.37	60.01	63.67	61.78	67.76	71.67	69.66
PARAPHRASE	62.99	58.30	60.55	70.87	70.90	70.89	60.80	64.98	62.82	70.35	74.04	72.15
EDWPN	56.38	48.78	53.76	72.84	65.30	68.82	63.58	55.83	60.25	68.35	67.28	68.03
SA-Transformer	61.28	48.98	54.44	70.76	65.85	68.22	62.82	58.31	60.48	72.01	62.87	67.13
ContrASTE	66.82	60.68	63.61	76.10	75.08	75.59	66.50	63.86	65.15	75.52	74.14	74.83
QAIE	\	\	38.58	\	\	\	\	\	43.82	\	\	51.41
LLaMA3-8B(zero-shot)	33.52	33.33	33.43	45.68	43.69	44.66	37.62	42.63	39.67	42.02	47.14	44.43
LLaMA3-8B(few-shot)	35.56	37.28	36.40	53.15	53.85	53.49	40.48	55.26	46.73	51.12	61.93	56.01
GPT-4o(zero-shot)	42.02	39.78	40.87	53.04	53.45	53.24	47.97	53.76	50.70	54.12	60.98	57.35
GPT-4o(few-shot)	42.55	46.94	44.63	61.12	66.23	63.57	53.00	60.57	56.53	58.68	72.04	64.68
DGSEP(S_1)	71.65	61.68	66.00	77.55	73.34	75.39	70.00	66.39	68.15	**75.98**	**75.10**	**75.54**
DGSEP(S_2)	**72.33**	63.77	**67.78**	77.51	73.84	75.63	**70.68**	66.60	**68.58**	75.40	73.93	74.66

average delay of the model during inference is about 46ms/sentence, and the maximum video memory occupancy is about 14GB, which is in line with the mainstream hardware environment and has good deployment feasibility.

4.4. Main results

This study uses precision (P), recall (R) and F1 value as evaluation indicators for the Aspect Sentiment Triplet Extraction (ASTE) task. By comparing the DGSEP model with four mainstream baseline methods, namely pipeline methods, end-to-end methods, two-stage methods and large language model-based methods, its performance advantage is verified. As shown in Table 2, on the four subsets of the V2 dataset, the DGSEP model generally surpasses the comparison models in the two core indicators of recall and F1, especially in the F1 dimension, showing a significant improvement: compared with the ContrASTE model, the 14lap, 14res, 15res and 16res subsets are improved by 4.17%, 0.04%, 3.43% and 0.71% respectively; compared with the SSGCN model, the improvement is 5.19%, 2.58%, 6.84% and 6.97% respectively. This performance gain may be attributed to two points: first, the simplified design based on the dual-stage prompt learning architecture effectively reduces the complexity of the task; second, the LSGF module strengthens the generation model's ability to parse complex language units through structured semantic modeling. Comprehensive experimental results show that the model can stably achieve significant superiority in the F1 indicator in cross-scenario data, verifying its technical effectiveness and methodological innovation in the ASTE task.

4.5. Model analysis

4.5.1. Ablation experiment

In this paper, we evaluate the contribution of the core components of the DGSEP model through systematic ablation experiments and control potential interference variables. As shown in Table 3, the F1 index analysis based on each subset of the V2 dataset shows that removing the label-oriented sequence label generation fusion module (LSGF) will lead to a significant decline in model performance, verifying the key role of this module in semantic structure modeling. At the same time, the two bar charts in Fig. 5 more clearly show the importance of this module; when the structured task prompt template T_1/T_2 is missing, the performance under the S_1 innovation strategy declines by 1.78%, 1.32%, 1.41% and 1.47% in subsets such as 14lap, respectively, revealing the irreplaceable role of the prompt template in task semantic guidance; while the removal of the Full component will lead to limited data diversity enhancement, resulting in 1.71%, 2.03%, 1.21% and 1.18% performance declines in each subset under the S_1 strategy. The experimental results fully prove that the components of the model work together in the triplet extraction process through a complementary mechanism, and its combined design is necessary to improve the robustness of the ASTE task.

Table 3
F1 scores of ablation study on V2.

Model	14lap	14res	15res	16res
DGSEP(S_1)	66.00	75.39	68.15	**75.54**
DGSEP(S_2)	**67.78**	**75.63**	**68.58**	74.66
W/0 LSGF(S1)	64.41	73.36	67.65	73.71
ΔF_1	-1.59	-2.03	-0.5	-0.83
W/0 LSGF(S2)	65.33	74.31	67.02	73.64
ΔF_1	-2.45	-1.32	-1.56	-1.02
W/0 T1(S1)	64.22	74.07	66.95	74.07
ΔF_1	-1.78	-1.32	-1.20	-0.47
W/0 T2(S2)	65.02	74.92	67.09	74.24
ΔF_1	-2.76	-0.71	-1.49	-0.42
W/0 Full (S1)	64.29	73.36	67.15	74.36
ΔF_1	-1.71	-2.03	-1.00	-0.18
W/0 Full (S2)	65.87	74.22	67.31	74.49
ΔF_1	-1.91	-1.41	-1.27	-0.17

4.5.2. Analysis of label sharing and boundary enforcement for entity splitting and merging

To further elucidate the principles behind the splitting and merging operations in LSGF, we conducted additional ablation experiments, focusing on sentences containing multi-word, overlapping, or nested

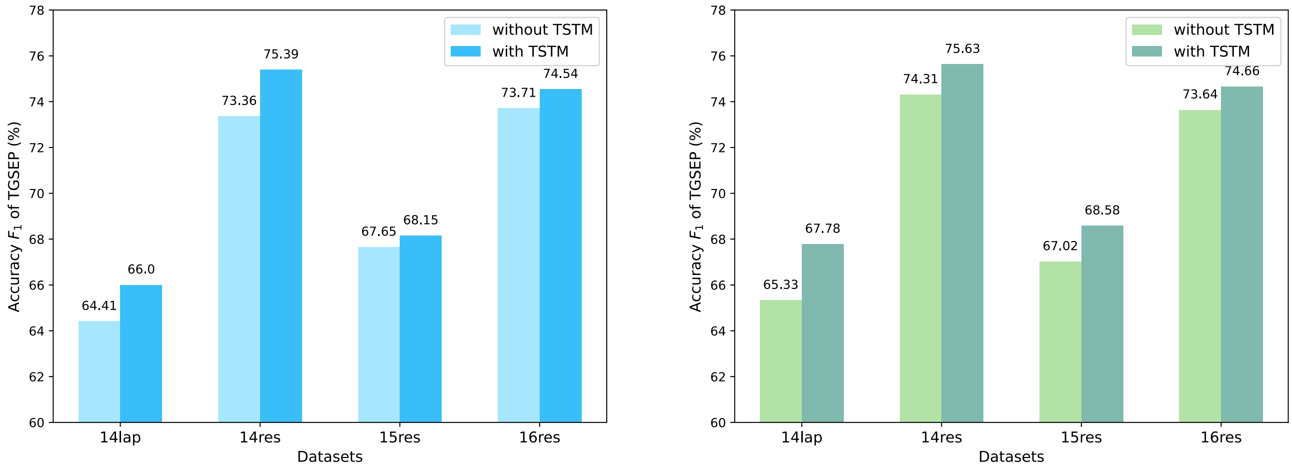

Fig. 5. F1 accuracy on the V2 dataset with and without the LSGF in DGSEP module.The left side uses innovation strategy S_1, and the right side uses innovation strategy S_2.

Table 4
Ablation study on boundary enforcement and label sharing for entity splitting and merging.

Model	Boundary F1	Splitting Accuracy	Merging Accuracy	Triplet F1
LSGF w/o Boundary Enforcement	82.3	74.5	88.1	78.4
LSGF w/o Label Sharing	90.1	79.2	83.5	81.2
Full LSGF (Ours)	**92.8**	**85.6**	**89.7**	**86.5**

Table 5
Controlled error propagation analysis.

Setting	Element F1	Triplet F1
Full DGSEP (ours)	**87.20**	**73.50**
+10% element drop	78.10	62.30
+10% element mislabeling	76.40	60.80
Oracle (gold elements input)	100.00	85.20

aspect/opinion terms. Specifically, we compared three variants: disabling boundary enforcement, allowing the model to freely split candidate spans without BIO constraints; disabling label sharing, preventing the same label sequence from being assigned to multiple entities; and including both boundary enforcement and label sharing. As shown in Table 4, boundary enforcement significantly improves span prediction consistency, resulting in a 10.5% improvement in boundary F1 and an 8.1% improvement in triplet F1 compared to the variant without boundary enforcement. It is shown that the splitting and merging in LSGF follows a decoding process governed by BIO rules. Label sharing can further reduce errors when multiple entities share the same label type, particularly in cases where entities such as “chicken” and “chicken soup” are both labeled with B-ASP. Without label sharing, these entities are more likely to be incorrectly merged. Enabling label sharing improves both merging accuracy and triplet F1, confirming that it helps the model correctly distinguish between different entities sharing the same label. Experimental results demonstrate that the full LSGF model (combining boundary enforcement and label sharing) achieves the best performance in terms of boundary F1, split accuracy, merge accuracy, and triplet F1. The experimental results show that the two mechanisms are complementary: boundary enforcement ensures span consistency, while label sharing maintains entity distinguishability, thus providing robustness and interpretability to the split/merge operations.

4.5.3. *Visual analysis of sequence labeling probability output*

The example in Fig. 6 shows the BIO probabilities output by the DGSEP model at each step. The figure above shows that the model made an error in the single element prediction stage and failed to correctly predict the aspect category; but in the second step of prediction, DGSEP successfully corrected this error and finally output the correct aspects and opinions, reflecting its fault tolerance and error correction capabilities. It shows that the innovative strategy we proposed plays a key role in processing structured complex semantics, and also verifies the feasibility of the dual-stage framework: by splitting the overall task into more fine-grained subtasks, it not only enhances the ability to capture implicit associations within tuples, but also improves the adaptability efficiency in language diversity scenarios, achieving a balance between flexibility and accuracy.

4.5.4. *Model complexity and running time analysis*

To better evaluate the computational efficiency and scalability of the proposed DGSEP framework, we analyze the model complexity and running time. The DGSEP framework adopts a two-stage architecture. Both stages are built on the T5 encoder-decoder model. Each stage contains 12 Transformer layers, the hidden layer size is $d = 768$, the feed-forward layer size is $d_{ff} = 3072$, and contains 12 attention heads. T represents the average input sequence length. In each stage, the computational complexity of a single Transformer block includes:

(1) Multi-head self-attention: $O(T^2 \cdot d)$
(2) Feed-forward network: $O(T \cdot d \cdot d_{ff})$

Each encoder and decoder has 12 layers, so the complexity of each stage is $O(L \cdot (T^2 \cdot d + T \cdot d \cdot d_{ff})) = O(12 \cdot (T^2 \cdot 768 + T \cdot 768 \cdot 3072))$. Since DGSEP adopts a two-stage design, the overall complexity becomes $O(2L \cdot (T^2 \cdot d + T \cdot d \cdot d_{ff}))$. Compared to single-stage models such as GAS or BART-ASTE, this two-stage pipeline introduces additional computational steps since both element-level and tuple-level predictions are performed sequentially. However, due to the modular nature of DGSEP, the two stages can be trained independently or in parallel, which improves overall flexibility and reduces training conflicts. On an NVIDIA RTX 3090 24GB GPU, the average inference time per sentence is about 46 milliseconds, slightly higher than the single-stage model, but still within a practically acceptable range. The moderate increase in latency is offset by a significant improvement in F1 scores across multiple benchmarks. In addition, since no additional heavy modules are introduced, the overall parameter size remains comparable to the T5-based baseline model, ensuring computational efficiency and deployment feasibility.

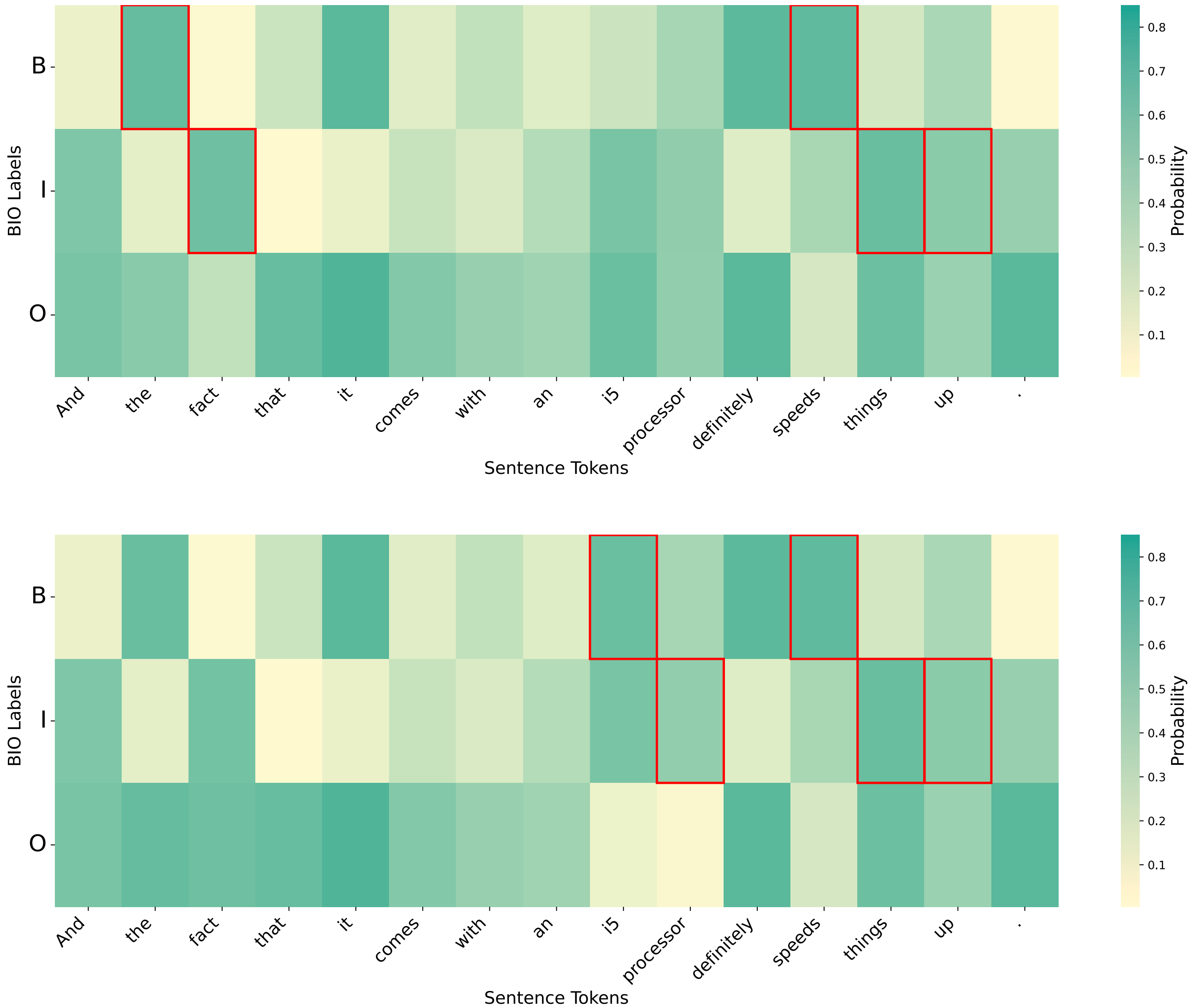

Fig. 6. Visualization of the sequence labeling (BIO) probability output of the DGSEP model. The above is the probability of the aspect and opinion tags predicted in the first step, and the following is the probability of the aspect and opinion tags predicted in the second step under the S_1 innovation strategy.

4.5.5. *Error propagation analysis*

In our two-stage framework, the generation of aspect sentiment triplets in the second stage depends to some extent on the quality of element predictions in the first stage. To explore potential error propagation, we conducted a controlled experiment by injecting noise into the output of the first stage–specifically, randomly removing or modifying 10% of the predicted elements. As shown in Table 5, this modification not only slightly reduces element-level accuracy, but also significantly lowers triplet-level F1 scores, highlighting the sensitivity of the second stage to upstream errors. Further case-level analysis shows that the impact of missing aspect or opinion terms is more detrimental than incorrect sentiment polarity classification. These findings highlight the importance of maximizing the recall of aspect and opinion terms in the first stage. In future work, we plan to mitigate error propagation by introducing a feedback mechanism between the two stages or combining a joint confidence scoring strategy.

4.5.6. *Case analysis*

Fig. 7 compares the performance of innovative strategies S_1 and S_2 in the ASTE task, taking two cases of the Laptop dataset as an example. Although S_1 had a bias in its judgment of the opinion category in the initial stage, S_2 missed the initial prediction of the aspect category, and two tuple-based prediction errors occurred in the second step, DGSEP still successfully output the correct result through the correction in the second step, reflecting the model's tolerance and error correction capabilities for prediction deviations. This phenomenon shows that the two strategies play a key supporting role in the model's processing of structured complex semantics. At the same time, the result verifies the feasibility of the dual-stage framework proposed in this paper: by decoupling the overall task into fine-grained subtasks, it not only enhances the model's ability to capture implicit associations within tuples, but also improves the adaptation efficiency of language diversity scenarios through a step-by-step processing mechanism, taking into account both flexibility and accuracy.

4.6. *Efficiency analysis*

To fully demonstrate the computational efficiency of DGSEP, we report inference latency and GPU memory usage on an NVIDIA RTX 3090 24GB GPU. Specifically, DGSEP takes an average of 46 milliseconds per sentence to infer, compared to 29.7 milliseconds for BART-ASTE and 36.5 milliseconds for GAS-T5. The increase in latency mainly stems from the two-stage generation process, where the second stage depends on output elements predicted by the first stage. In terms of memory usage,

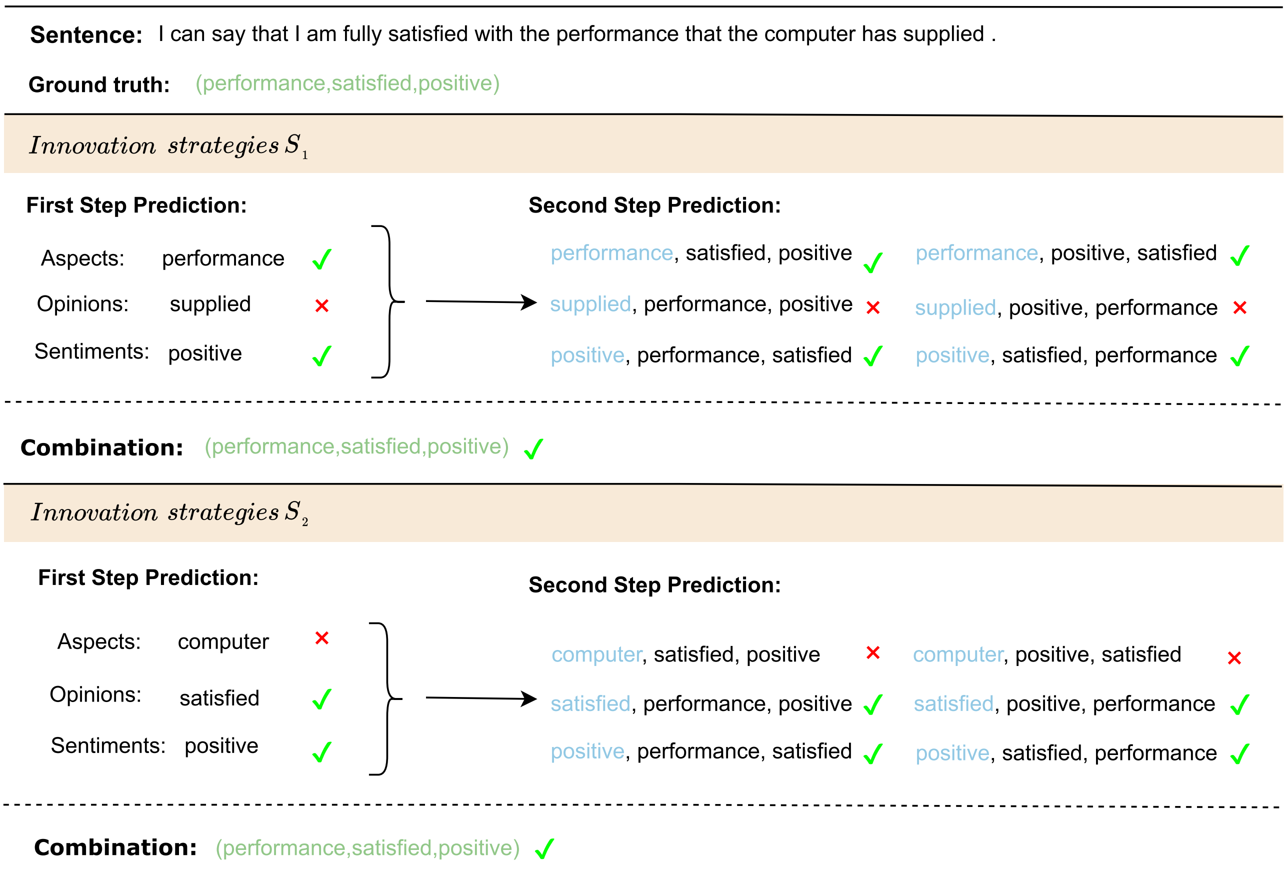

Fig. 7. Example of the ASTE task. Aspects and their corresponding opinions are highlighted using the same color.

DGSEP requires approximately 14GB of GPU memory during inference, which is comparable to other T5-based models and still within the capacity of mainstream hardware. Despite the slightly higher computational cost, DGSEP consistently maintains excellent F1 performance on all datasets, achieving a good balance between accuracy and efficiency.

5. Conclusion

We propose DGSEP, a new dual-stage prompting method that effectively exploits data augmentation by using single elements for tuple prediction, while decomposing the task into smaller subtasks, improving flexibility and efficiency, better capturing dependencies within tuples, and adapting to language diversity. DGSEP provides two innovative strategies aimed at achieving different prompt template styles and creating a data-efficient method that can make predictions with significantly reduced data augmentation rates. It also proposes a label-oriented sequence label generation fusion module (LSGF), which aims to fuse T5-based and label-oriented sequence labels to improve the ability of the generation model to handle complex structures. DGSEP outperforms most baselines in aspect-level sentiment triplet extraction (ASTE), demonstrating its potential to improve efficiency and accuracy of NLP tasks.

Although our proposed strategy achieves state-of-the-art performance, it also has some limitations. The strict enforcement of BIO boundaries can also introduce limitations in sentences containing overlapping or nested entities. For example, in the sentence “spicy chicken soup”, when both “chicken” and “chicken soup” are valid aspect terms, enforcing consistency may incorrectly merge them into one span or inappropriately split them. Although this situation is relatively rare in the benchmark dataset, we emphasize this as a potential limitation of our approach. To alleviate this issue, future work will extend LSGF to use span-level attention mechanisms or dynamic boundary constraints to more flexibly handle partially overlapping spans.

At the same time, a filtering mechanism is introduced to minimize errors by preventing potentially mispredicted single elements from propagating between steps, which is crucial to maintaining the integrity of the results. In addition, since it is based on a dual-stage hint architecture, improvements in any step or subtask will improve the overall performance, so we should focus on improving each step and training more expert models.

CRediT authorship contribution statement

Yujun Chen: Conceptualization, Methodology, Writing - original draft; **Mingwei Tang:** Funding acquisition, Writing - review & editing, Data curation; **Shangyi Du:** Visualization, Validation; **Kun Yang:** Formal analysis, Validation; **Yanxi Zheng:** Supervision, Resources; **Mingfeng Zhao:** Formal analysis.

Data availability

Data will be made available on request.

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.

Acknowledgments

This work is supported by Sichuan Science and Technology Program (Nos: 2025NSFSC2017, 25LHJJ0133), the Scientic Research

Funds project of Science and Technology Department of Sichuan Province (Nos: 2019YFG0508, 2019GFW131, 24NSFJQ0015), Sichuan Key R&D project (No. 2023YFG0354), the National Natural Science Foundation of China (No. 61902324), Funds Project of Chengdu Science and Technology Bureau (Nos. 2017-RK00-00026-ZF, 2022-YF04-00065-JH, 2023-JB00-00020-GX), the National Natural Science Foundation of China (No. 62576295)

References

Bao, X., Wang, Z., Jiang, X., Xiao, R., & Li, S. (2022). Aspect-based sentiment analysis with opinion tree generation. In *Ijcai* (pp. 4044–4050). (*vol. 2022*).

Cai, H., Tu, Y., Zhou, X., Yu, J., & Xia, R. (2020). Aspect-category based sentiment analysis with hierarchical graph convolutional network. In *Proceedings of the 28th international conference on computational linguistics* (pp. 833–843).

Chen, C., Teng, Z., & Zhang, Y. (2020). Inducing target-specific latent structures for aspect sentiment classification. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 5596–5607).

Chen, H., Zhai, Z., Feng, F., Li, R., & Wang, X. (2022). Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 1: long papers)* (pp. 2974–2985).

Chen, S., Wang, Y., Liu, J., & Wang, Y. (2021). Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 12666–12674). (*vol. 35*).

Dai, D., Chen, T., Xia, S., Wang, G., & Chen, Z. (2022). Double embedding and bidirectional sentiment dependence detector for aspect sentiment triplet extraction. *Knowledge-Based Systems*, *253*, 109506.

Fei, H., Ren, Y., Zhang, Y., & Ji, D. (2021). Nonautoregressive encoder–decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Transactions on Neural Networks and Learning Systems*, *34*, 5544–5556.

Gao, C., Zhang, X., Li, L., Li, J., Zhu, R., Du, K., & Ma, Q. (2023). Ergm: A multi-stage joint entity and relation extraction with global entity match. *Knowledge-Based Systems*, *271*, 110550.

Gou, Z., Guo, Q., & Yang, Y. (2023). Mvp: Multi-view prompting improves aspect sentiment tuple prediction. *arXiv preprint arXiv:2305.12627*.

Hao, S., Zhou, Y., Liu, P., & Xu, S. (2024). Bi-syntax guided transformer network for aspect sentiment triplet extraction. *Neurocomputing*, *594*, 127880.

Huang, L., Wang, P., Li, S., Liu, T., Zhang, X., Cheng, Z., Yin, D., & Wang, H. (2021). First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction. *arXiv preprint arXiv:2102.08549*.

Li, P., Li, P., & Zhang, K. (2023a). Dual-channel span for aspect sentiment triplet extraction. In *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 248–261).

Li, Q., Wen, W., & Qin, J. (2024). Improving span-based aspect sentiment triplet extraction with part-of-speech filtering and contrastive learning. *Neural Networks*, *177*, 106381.

Li, X., Li, D., Du, R., Chen, D., & Madden, A. (2023b). Double policy network for aspect sentiment triplet extraction (student abstract). In *Proceedings of the AAAI conference on artificial intelligence* (pp. 16256–16257). (*vol. 37*).

Li, Y., He, Q., Du, N., & He, Q. (2025). Implicit expression recognition enhanced table-filling for aspect sentiment triplet extraction. *Neurocomputing*, *614*, 128776.

Li, Y., Lin, Y., Lin, Y., Chang, L., & Zhang, H. (2022). A span-sharing joint extraction framework for harvesting aspect sentiment triplets. *Knowledge-Based Systems*, *242*, 108366.

Liu, P., Joty, S., & Meng, H. (2015). Fine-grained opinion mining with recurrent neural networks and word embeddings. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1433–1443).

Liu, Y., Zhou, Y., Li, Z., Wang, J., Zhou, W., & Hu, S. (2023). Him: An end-to-end hierarchical interaction model for aspect sentiment triplet extraction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *31*, 2272–2285.

Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Lu, H.-y., Liu, T.-c., Cong, R., Yang, J., Gan, Q., Fang, W., & Wu, X.-j. (2025). Qaie: Llm-based quantity augmentation and information enhancement for few-shot aspect-based sentiment analysis. *Information Processing & Management*, *62*, 103917.

Lu, Q., Sun, X., Sutcliffe, R., Xing, Y., & Zhang, H. (2022). Sentiment interaction and multi-graph perception with graph convolutional networks for aspect-based sentiment analysis. *Knowledge-Based Systems*, *256*, 109840.

Lu, Y., Lin, H., Xu, J., Han, X., Tang, J., Li, A., Sun, L., Liao, M., & Chen, S. (2021). Text2event: Controllable sequence-to-structure generation for end-to-end event extraction. *arXiv preprint arXiv:2106.09232*.

Ma, D., Li, S., Wu, F., Xie, X., & Wang, H. (2019). Exploring sequence-to-sequence learning in aspect term extraction. In *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 3538–3547).

Mao, Y., Shen, Y., Yu, C., & Cai, L. (2021). A joint training dual-mrc framework for aspect based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 13543–13551). (*vol. 35*).

Peng, H., Xu, L., Bing, L., Huang, F., Lu, W., & Si, L. (2020). Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 8600–8607). (*vol. 34*).

Shi, L., Han, D., Han, J., Qiao, B., & Wu, G. (2022). Dependency graph enhanced interactive attention network for aspect sentiment triplet extraction. *Neurocomputing*, *507*, 315–324.

Simmering, P. F., & Huoviala, P. (2023). Large language models for aspect-based sentiment analysis. *arXiv preprint arXiv:2310.18025*.

Sun, Q., Yang, L., Ma, M., Ye, N., & Gu, Q. (2024). Rethinking aste: A minimalist tagging scheme alongside contrastive learning. *arXiv preprint arXiv:2403.07342*.

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 *Proceedings of the AAAI conference on artificial intelligence* (pp. 9122–9129). (*vol. 34*).

Wang, W., Pan, S. J., Dahlmeier, D., & Xiao, X. (2017). Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In *Proceedings of the AAAI conference on artificial intelligence*. (*vol. 31*).

Wu, M., Wang, W., & Pan, S. J. (2020a). Deep weighted maxsat for aspect-based opinion extraction. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 5618–5628).

Wu, S., Fei, H., Ren, Y., Li, B., Li, F., & Ji, D. (2021). High-order pair-wise aspect and opinion terms extraction with edge-enhanced syntactic graph convolution. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *29*, 2396–2406.

Wu, Z., Ying, C., Zhao, F., Fan, Z., Dai, X., & Xia, R. (2020b). Grid tagging scheme for aspect-oriented fine-grained opinion extraction. *arXiv preprint arXiv:2010.04640*.

Xia, T., Sun, X., Yang, Y., Long, Y., & Sutcliffe, R. (2024). A dual relation-encoder network for aspect sentiment triplet extraction. *Neurocomputing*, *597*, 128064.

Xu, G., Yang, Z., Xu, B., Luo, L., & Lin, H. (2025a). Span-based syntactic feature fusion for aspect sentiment triplet extraction. *Information Fusion*, *120*, 103078.

Xu, H., Tang, M., Cai, T., Hu, J., Jiang, Z., Bian, D., & Lv, S. (2025b). Multiple-level enhanced graph convolutional network for aspect sentiment triplet extraction. *NEUROCOMPUTING*, *634*. https://doi.org/10.1016/j.neucom.2025.129834

Xu, H., Tang, M., Cai, T., Hu, J., & Zhao, M. (2024a). Dual-enhanced generative model with graph attention network and contrastive learning for aspect sentiment triplet extraction. *KNOWLEDGE-BASED SYSTEMS*, *301*. https://doi.org/10.1016/j.knosys.2024.112342

Xu, H., Zhang, D., Zhang, Y., & Xu, R. (2024b). Hitsz-hlt at sighan-2024 dimabsa task: Integrating bert and llm for chinese dimensional aspect-based sentiment analysis. In *Proceedings of the 10th SIGHAN workshop on chinese language processing (SIGHAN-10)* (pp. 175–185).

Xu, L., Chia, Y. K., & Bing, L. (2021). Learning span-level interactions for aspect sentiment triplet extraction. *arXiv preprint arXiv:2107.12214*.

Xu, L., Li, H., Lu, W., & Bing, L. (2020). Position-aware tagging for aspect sentiment triplet extraction. *arXiv preprint arXiv:2010.02609*.

Yan, H., Dai, J., Qiu, X., Zhang, Z. et al. (2021). A unified generative framework for aspect-based sentiment analysis. *arXiv preprint arXiv:2106.04300*.

Yang, K., Zong, L., Tang, M., Hu, J., Zheng, Y., Chen, Y., & Zhao, M. (2025a). Mpgm:multi-prompt generation model with self-supervised contrastive learning for aspect sentiment triplet extraction. *NEURAL NETWORKS*, *192*. https://doi.org/10.1016/j.neunet.2025.107894

Yang, K., Zong, L., Tang, M., Zheng, Y., Chen, Y., Zhao, M., & Jiang, Z. (2025b). Mpbe: Multi-perspective boundary enhancement network for aspect sentiment triplet extraction. *APPLIED INTELLIGENCE*, *55*(4). https://doi.org/10.1007/s10489-024-06144-z

Yang, S., Zhang, T., Xu, H., & Jia, Y. (2023a). Improving aspect sentiment triplet extraction with perturbed masking and edge-enhanced sentiment graph attention network. In *2023 international joint conference on neural networks (IJCNN)* (pp. 1–8). IEEE.

Yang, X., Peng, T., Bi, H., & Han, J. (2024). Span-level bidirectional retention scheme for aspect sentiment triplet extraction. *Information Processing & Management*, *61*, 103823.

Yang, Y., Sun, X., Lu, Q., Sutcliffe, R., & Feng, J. (2023b). A sentiment and syntactic-aware graph convolutional network for aspect-level sentiment classification. In *Icassp 2023-2023 ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 1–5). IEEE.

Yang, Y., Zhou, S., & Liu, Y. (2023c). Bidirectional relation-guided attention network with semantics and knowledge for relational triple extraction. *Expert Systems with Applications*, *224*, 119905.

Yu, J., Jiang, J., & Xia, R. (2018). Global inference for aspect and opinion terms co-extraction based on multi-task neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *27*, 168–177.

Yuan, L., Wang, J., Yu, L.-C., & Zhang, X. (2023). Encoding syntactic information into transformers for aspect-based sentiment triplet extraction. *IEEE Transactions on Affective Computing*, *15*, 722–735.

Zhang, C., Li, Q., Song, D., & Wang, B. (2020). A multi-task learning framework for opinion triplet extraction. *arXiv preprint arXiv:2010.01512*.

Zhang, H., Cheah, Y.-N., Alyasiri, O. M., & An, J. (2024). Exploring aspect-based sentiment quadruple extraction with implicit aspects, opinions, and chatGPT: a comprehensive survey. *Artificial Intelligence Review*, *57*, 17.

Zhang, J., Xu, S., Gao, X., & Tang, Z. (2025). Aspect sentiment triplet extraction with syntax-semantics graph convolutional network. *International Journal of Computational Intelligence Systems*, *18*, 167.

Zhang, W., Deng, Y., Li, X., Bing, L., & Lam, W. (2021a). Aspect-based sentiment analysis in question answering forums. In *Findings of the association for computational linguistics: EMNLP 2021* (pp. 4582–4591).

Zhang, W., Deng, Y., Li, X., Yuan, Y., Bing, L., & Lam, W. (2021b). Aspect sentiment quad prediction as paraphrase generation. *arXiv preprint arXiv:2110.00796*.

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2021c). Towards generative aspect-based sentiment analysis. Association for Computational Linguistics.

Zhang, W., Li, X., Deng, Y., Bing, L., & Lam, W. (2022a). A survey on aspect-based sentiment analysis: Tasks, methods, and challenges. *IEEE Transactions on Knowledge and Data Engineering*, *35*, 11019–11038.

Zhang, Y., Yang, Y., Li, Y., Liang, B., Chen, S., Dang, Y., Yang, M., & Xu, R. (2022b). Boundary-driven table-filling for aspect sentiment triplet extraction. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 6485–6498).

Zhao, H., Huang, L., Zhang, R., Lu, Q. et al. (2020). Spanmlt: A span-based multi-task learning framework for pair-wise aspect and opinion terms extraction. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 3239–3248).