Article

Duality-Driven Aspect Sentiment Triplet Extraction with LLM and Iterative Reinforcement

Xun Li, Kun Zhang * and Danjie Han

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China; 219106010015@njust.edu.cn (X.L.); handanjie@njust.edu.cn (D.H.)
* Correspondence: zhangkun@njust.edu.cn

Abstract: Aspect-based sentiment triplet extraction tasks remain a long-standing challenge, which aim to achieve aspect, opinion, and sentiment polarity from sentences. Most existing methods achieve excellent performance by exploring the interactions between aspect and opinion terms. However, few studies focus on the positive impact of sentiment on triplet extraction. As sentiment acts as a key cue in triplet extraction, its role is often overlooked, thereby limiting extraction performance. This paper proposes a novel framework, duality-driven aspect sentiment triplet extraction with a large language model and iterative reinforcement, which integrates duality-driven with a large language model for the aspect sentiment triplet task. This study employs a duality-driven strategy based on symmetry to extract aspect-based sentiment triplets, fully taking into account sentiment polarity during the interaction between aspects and opinions. Moreover, this study devises a two-view prompt template for prior knowledge fusion based on large language models and employs confidence cycle iteration strategies to alleviate cascading errors. Extensive experiments show that the framework outperforms the previous state-of-the-art model. These findings demonstrate that the proposed model makes a positive impact on the aspect sentiment triplet extraction task overall.

Keywords: aspect sentiment triplet extraction; duality-driven extraction; large language model

1. Introduction

With the rapid development of e-commerce, users have shared vast amounts of content on various platforms, and this content is rich in valuable information. For manufacturers, customer reviews encapsulate critical consumer opinions and sentiment trends regarding their products, providing essential feedback that can inform product improvements and optimization strategies. Conversely, for consumers, detailed reviews facilitate a rigorous evaluation of product quality, empowering them to make well-informed decisions that align with their specific needs. Consequently, employing effective techniques to extract valuable information from reviews is of paramount importance. Aspect-based sentiment analysis (ABSA) techniques, which are capable of identifying aspect-specific keywords, sentiment polarities, and opinion orientations, have been widely adopted across diverse sectors, including restaurants, electronics, apparel, and automobiles, to robustly support the extraction and analysis of user feedback [1–3].

In the field of sentiment analysis, ABSA plays a pivotal role by systematically capturing three key elements: aspect terms, opinion terms, and sentiment polarity [4]. Building on these elements, ABSA encompasses a range of subtasks, including aspect term extraction (AE), opinion term extraction (OE), aspect extraction with sentiment classification (AESC),

and aspect and opinion extraction (AOE), culminating in the more challenging aspect–opinion–sentiment triplet extraction (ASTE). To address the challenge of ASTE, numerous studies have proposed solutions that can be broadly categorized into two approaches: two-stage methods and joint models. In the two-stage paradigm, separate sequence labeling models are first employed to independently extract potential aspects and opinions from a sentence, after which sentiment polarity is predicted based solely on the extracted aspects [5]. However, because these stages are trained independently and do not share information during triplet extraction, overall performance is often compromised. Consequently, many researchers have shifted towards joint models that integrate information across all stages, thereby enabling the simultaneous extraction of aspects, opinions, and sentiment [6–8]. Nonetheless, even joint extraction methods typically follow a sequential strategy, first extracting aspect and opinion terms and then predicting sentiment polarity based on the extracted elements [9,10]. These approaches primarily model the interaction between aspects and opinions while largely overlooking the reinforcing influence of sentiment on this interaction. For example, when a sentence conveys a positive sentiment, the model is more likely to interpret the evaluation of a vase as positive rather than negative. Despite these advances, current research indicates that the performance of sentiment triplet extraction remains limited.

Liu et al. [11] introduced a duality-driven learning model based on sequential prompting for ASTE. This model employs a bidirectional extraction framework that integrates sentiment polarity into the interaction between aspects and opinions. In one direction, the model sequentially extracts the aspect, followed by sentiment and then the opinion; in the reverse direction, it extracts the opinion first, followed by the aspect and sentiment. Although this approach yields performance improvements in triplet extraction, it suffers from a critical limitation: the sequential dependency means that errors in early extraction stages (e.g., aspect extraction) may propagate and amplify in subsequent steps. To overcome these shortcomings, we propose a large language model (LLM) and iterative enhanced Dual-direction ASTE method (DASTER). Large language models (LLMs), trained on vast corpora, exhibit remarkable natural language understanding capabilities, enabling them to capture deep semantic representations and intricate contextual relationships with high precision. Our method combines a LLM with a two-view prompt template, augmented by multi-round iterative refinement to mitigate cascading errors. Our framework initially employs an LLM with two-view prompts to generate candidate triplets (i.e., aspect, opinion, and sentiment polarity), which are subsequently encoded into feature vectors using bidirectional encoder representations from transformers (BERT). We then concatenate the semantic priors from the LLM with the BERT-encoded sentence features, yielding a rich representation infused with triplet information. This enhanced representation is input into two extraction models corresponding to the two extraction directions, namely, the Aspect-to-Sentiment-and-Opinion (A2SO) model and the Opinion-to-Aspect-and-Sentiment (O2AS) model. Since the triplets derived from these two directions should theoretically be consistent, we introduce a Kullback–Leibler (KL) divergence loss to measure and minimize the distance between the respective probability distributions.

Furthermore, in the triplet extraction phase, we incorporate a multi-round iterative algorithm. In each iteration, the model identifies and retains high-confidence triplets while re-predicting those with lower confidence. This iterative refinement progressively reduces the impact of cascading errors, thereby significantly enhancing extraction accuracy and overall robustness.

Our main contributions are summarized as follows:

1. This paper proposes a novel framework that extends the dual-driven paradigm by integrating large language models with a multi-round iterative enhancement

    mechanism, thereby mitigating the cascading errors inherent in sequential extraction for aspect-based sentiment triplet extraction.

2.   We propose an innovative fusion technique that combines prior knowledge of LLM with the source sentence via a multi-head attention mechanism, thereby supplying effective triplet priors to the dual-driven module. Unlike previous work, we design prompting templates for two distinct views, sentiment and semantic, and employ outputs of LLMs from each perspective to extract aspect-based triplets.

3.   The experimental results demonstrate the robustness of our proposed framework. Furthermore, a comprehensive evaluation was performed to analyze the contribution of each key module. The analysis further demonstrates the strong capability of framework in extracting aspect-based sentiment triples.

2. Related Work

In this section, we present a systematic review of the related work in ASTE. Existing studies in this field can be broadly classified into two categories: two-stage and joint extraction approaches. To provide a comprehensive overview of the current research landscape, we discuss these two methodologies in detail.

2.1. Two-Stage Methods

Two-stage methods decompose the overall task into two sequential subtasks: the first stage focuses on information extraction, for example, identifying and extracting aspect terms and opinion terms from the text; the second stage classifies the sentiment polarity, positive, negative, or neutral, for the candidate aspect–opinion pairs generated in the first stage [7]. By decoupling extraction from classification, this approach simplifies the model design and enables the application of specialized techniques to each subtask. However, the independent training of these stages can lead to error propagation, which may ultimately impair overall performance.

Peng et al. [5] were the first to propose the aspect-based sentiment triplet extraction (ASTE) task and designed a two-stage approach to address it. In the first stage, two distinct sequence labeling models are employed to extract aspect terms that contain sentiment information as well as opinion terms; in the second stage, the extracted aspects and opinions are combined, and a classifier is constructed to determine the validity of the pairings. Subsequently, Chen et al. [10] and Mao et al. [9] defined the ASTE task as a multi-step machine reading comprehension (MRC) problem and developed a corresponding framework. Li et al. [12] further advanced the field by introducing a dual-strategy network, DPM-MTR, in which an upper-level tagging network extracts all aspects from the input sample, and a lower-level network subsequently extracts the corresponding opinions and classifies sentiment based on aspect-specific representations. Similarly, Li et al. [12] devised a three-level sequence labeling strategy that sequentially extracts aspects, then opinions, and finally classifies sentiment, again relying on aspect-specific representations. Despite these advances, the performance of two-stage methods is often undermined by error propagation between independently trained stages, procedural complexity, and a failure to fully exploit the interdependencies among aspects, opinions, and sentiment during extraction [13].

2.2. Joint Extraction Methods

Joint extraction methods simultaneously extract aspect terms, opinion terms, and sentiment polarity within a unified model, thereby eliminating the need to decompose the task into separate steps [11]. By leveraging end-to-end training, these methods fully exploit the intrinsic relationships among various textual elements, enabling coordinated

information extraction. This unified strategy not only mitigates the error propagation issues that are common in pipeline approaches but also, to some extent, enhances the overall accuracy and robustness of the system.

Xu et al. [6] introduced a position-aware sequence labeling method that effectively models the rich interdependencies among sentiment elements, while Yan et al. [14] combined the generative power of BART with a pointer network to precisely locate target expressions. Zhang et al. [15] further contributed by proposing two paradigms annotation style and extraction style within a unified generative framework to address multi-sentiment analysis tasks.

To enhance performance further, Yuan et al. [16] augmented Transformer blocks with a syntax-aware layer, embedding syntactic information into word-pair representations. Complementing these efforts, Zhang et al. [17] proposed a multi-task learning framework that employs a shared encoder to simultaneously extract the components of sentiment triplets via three subtasks. Liu et al. [18] adopted a bidirectional MRC framework, where one strategy sequentially extracts aspect terms followed by opinion terms, and an alternative strategy reverses this order. Jing et al. [8] introduced a dual-encoder architecture to explicitly model differences between subtasks, while Chen et al. [19] designed a multi-channel graph convolutional network that leverages semantic dependency trees and diverse graph channels to construct robust word relationships. Moreover, Zhang et al. [20] proposed a boundary detection model that integrates interactions between lexical items and their relational context. In another innovative contribution, Shi et al. [21] employed a dual affine attention module to extract targeted semantic information by defining four linguistic relationships for word pairs, thereby enhancing word dependencies and subsequently integrating features via a graph convolutional network. Finally, Li et al. [22] developed a dual graph convolutional network (dual-GCN) model that concurrently integrates SenticNet sentiment knowledge with positional information, further advancing the state of the art in this domain.

Although the methods described above consider the interactions between aspects and opinions, they generally overlook the beneficial role of sentiment information within the triplet structure. For instance, Peng et al. [23] treat original table cells as nodes and compute edge weights using prompt attention scores to construct a target-aware grid graph. Yang et al. [24] introduced a bidirectional retention strategy to counteract semantic dilution between aspect and opinion terms caused by their separation. Zou et al. [25] further advanced the field by integrating multi-task shared cascade learning with MRC, thereby proposing a triplet MRC framework that balances the contributions of its components. Additionally, Liu et al. [11] formulated a bidirectional extraction binary equation to extract triplets from different directions and impose constraints on them. Nonetheless, these approaches either fail to fully exploit the mutual dependencies among triplet elements or suffer from error propagation, rendering aspect-based sentiment triplet extraction a persistently challenging problem.

Based on the above research, this paper introduces a novel framework that extends the duality-driven paradigm by integrating LLM with a multi-round iterative reinforcement mechanism, thereby mitigating the inherent cascading errors in aspect-based sentiment triplet extraction. Our proposed approach significantly differs from multi-phase querying methods such as SynPrompt [4], which constructs prompts based solely on English syntactic knowledge, and Dual-MRC [9], which sequentially extracts aspect terms, opinions, and sentiment polarities via dual machine reading comprehension models. Specifically, our framework employs a dual-view prompting strategy that independently extracts aspect-based sentiment triplets from semantic and sentiment perspectives, facilitating bidirectional correction between semantic and sentiment-driven outputs.

3. The Proposed Method

In this section, we first define the research problem, clarifying our study objectives and overall direction. We then provide a systematic description of the model architecture, emphasizing its design principles and operational mechanisms. Finally, we offer a detailed analysis of each component, describing their functions and implementations, with the goal of presenting a comprehensive exposition of the theoretical and practical framework that constructs this study.

3.1. Problem Definition

Figure 1 shows an example of ABSA task. For example, consider the sentence "This vase is exquisite, but the price is expensive". This sentence contains two aspect terms, "vase" and "price", each associated with distinct sentiment polarities: the sentiment toward "vase" is positive, whereas the sentiment toward "price" is negative. This paper focuses on the ASTE task, which aims to extract triplets, comprising an aspect, an opinion, and a sentiment polarity, from a sentence. In the given example, two triplets can be extracted: (vase, exquisite, and positive) and (price, expensive, and negative). Compared with other ABSA subtasks, ASTE is more challenging because it requires the model not only to accurately identify all the relevant elements (i.e., aspects, opinions, and sentiment polarities) within a sentence but also to effectively leverage the interrelations among these components to achieve a more precise extraction.
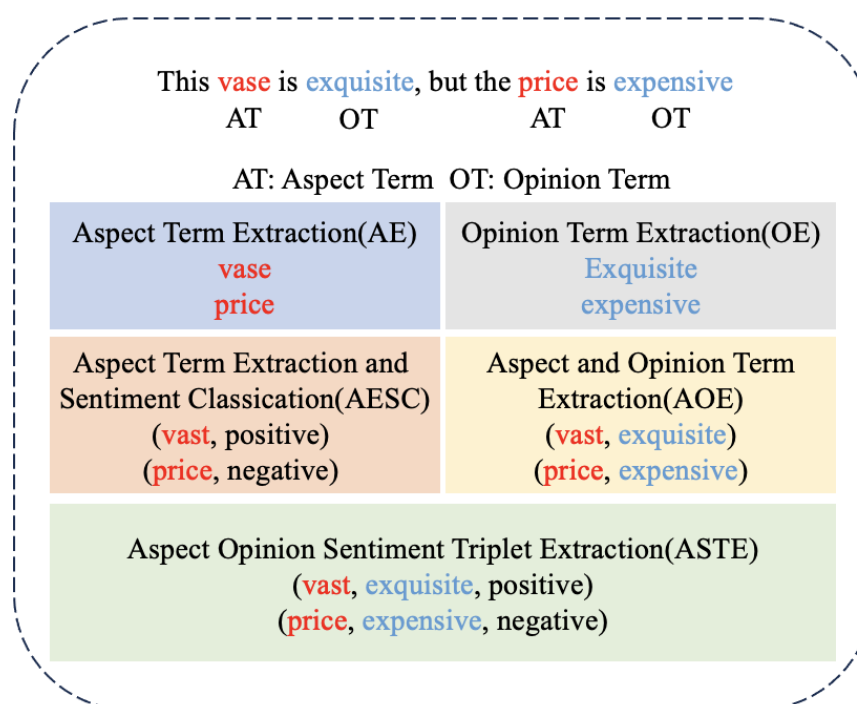

Figure 1. An example of ABSA task.

Given a sentence $W = [w_1, w_2, \ldots, w_n]$ that contains n words, and a triplet set $T = \{(a_i, o_i, s_i)\}_{i=1}^{|T|}$, where each triplet consists of an aspect term a_i, an opinion term o_i and a sentiment polarity s_i with $|T|$ denoting the total number of triplets, the objective of the aspect-based sentiment triplet extraction task is to extract the aspect terms, opinion terms, and corresponding sentiment polarities from a given sentence W to form triplets. Both aspect and opinion terms can be either a single word or a phrase within the sentence, and the sentiment polarity is categorized as positive, neutral, or negative. For example, in the sentence "This vase is very beautiful, but the price is somewhat expensive", the aspect–

opinion–sentiment triplets are (vase, beautiful, and positive) and (price, expensive, and negative). In the aspect-based sentiment triplet extraction task, different triplets extracted from the same sentence are treated as independent entities. For a given sentence W, the extracted triplet (a, o, s) can be formalized as follows:

$$argmax\, p(a, o, s|W) \tag{1}$$

where function $argmax$ is the function of selecting the most probable aspect–opinion–sentiment triplet (a, o, s) from a given sentence W.

3.2. The Proposed Framework Structure

In this paper, we propose a model that enhances the binary aspect-based sentiment triplet extraction framework by integrating a large language model (LLM), multi-round iterative extraction, and KL divergence regularization. As illustrated in Figure 2, our approach begins by employing an LLM of two-view prompt template to generate initial candidate triplets, comprising aspects, opinions, and sentiment, from the input sentence. These candidate triplets are then used to enrich the contextual representation of sentence by fusing the LLM-derived embeddings with BERT-based encodings via a multi-head attention mechanism. This enhanced context forms the foundation for our multi-round iterative extraction process. In the bidirection triplet extraction phase, two extraction directions are employed: one proceeds from aspect to sentiment to opinion (A2SO), and the other follows the order of opinion to aspect to sentiment (O2AS). By leveraging these dual extraction strategies, our model is capable of extracting aspect–opinion–sentiment triplets from sentences in a more comprehensive and accurate manner.
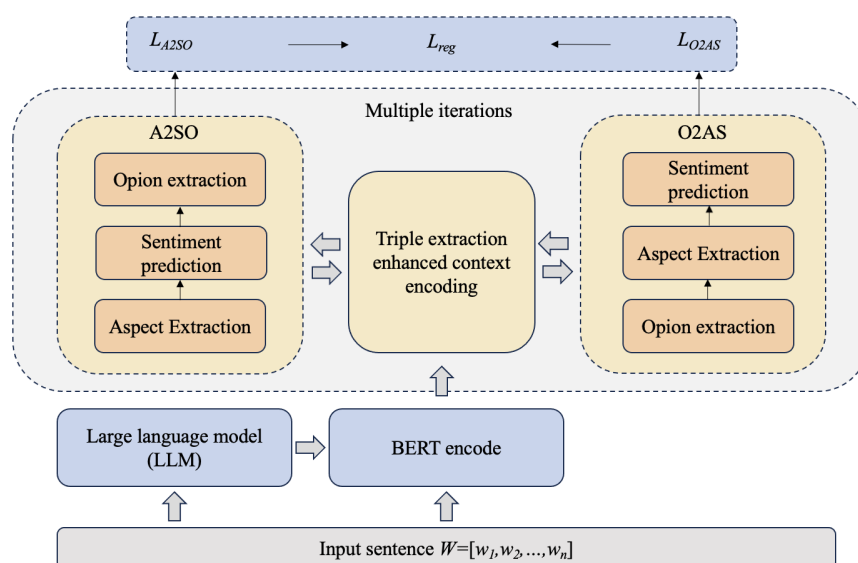

Figure 2. The framework of DASTER.

The A2SO method operates in three stages. First, aspect terms are extracted from the input sentence. Based on these extracted aspects, the corresponding sentiment polarity and opinion terms are subsequently identified. Formally, this process can be defined as follows:

$$p_A(a, o, s|W) = p(a|W) \cdot p(s|a, W) \cdot p(o|a, s, W) \tag{2}$$

The O2SA method first extracts opinion terms from the sentence; based on these, it further identifies the corresponding aspect terms and sentiment polarity. Specifically, this process can be formalized as follows:

$$p_O(a, o, s|W) = p(o|W) \cdot p(a|o, W) \cdot p(s|a, o, W) \tag{3}$$

where $p(a|W)$ and $p(o|W)$ are produced by the Global Pointer Network (GPN) [26] followed by a softmax layer, whereas $p(s|\cdot)$ is generated by a word-level soft-mask layer, each conditional term is produced by a softmax layer. Thus, all these are proper probability distributions.

To mitigate cascading errors arising from reliance on previous outputs, we employ a confidence evaluation method to iteratively refine the model. During multiple iterations, the extraction of triplets is progressively optimized: in each iteration, high-confidence triplets are retained to update the contextual information, while low-confidence triplets are re-extracted, thereby improving extraction accuracy and result consistency.

Since the triplets extracted by the A2SO and O2AS methods are required to be consistent, we introduce a KL divergence regularization loss based on the correlation between their probability distributions. Specifically, the KL divergence regularization loss is computed as follows:

$$Loss_{reg}^{(k)} = \frac{1}{2}(D_{KL}(p_A^{(k)}(a, o, s|W)||p_O^{(k)}(a, o, s|W)) + D_{KL}(p_O^{(k)}(a, o, s|W)||p_A^{(k)}(a, o, s|W))) \tag{4}$$

where k represents the iteration count, and D_{KL} represents the KL divergence between the probability distributions of triplets extracted by the A2SO and O2AS methods.

3.3. Triple Generation Based on Large Language Model

In our approach, we first employ a LLM to extract aspect-based sentiment triplets, which serve as prior knowledge for the duality module. To ensure the reliability of this prior knowledge, we propose a dual-view prompt strategy that independently extracts triplets from both semantic and sentiment perspectives. By leveraging the LLM as an initial reference, our iterative refinement process effectively reduces bias in the initial predictions.

To mitigate semantic ambiguity and achieve fine-grained sentiment detection in aspect-based sentiment extraction, we introduce a novel method based on a dual-view prompt template. Our approach harnesses the powerful semantic understanding and generative capabilities of LLMs to simultaneously extract aspect-based sentiment triplets from both the semantic context and explicit sentiment cues. This dual-view strategy enhances extraction accuracy and robustness by effectively bridging the gap between implicit contextual meanings and overt sentiment expressions.

3.3.1. Semantic Prompt Template

This template prioritizes comprehensive semantic understanding, requiring the model to accurately identify specific aspects, their associated opinion descriptions, and the sentiment polarity deduced from the context of the input sentence. Details are illustrated in the Figure 3.

Semantic prompt template:	Sentiment prompt template:
Please read the following text carefully and use your semantic comprehension to identify and extract aspect-based sentiment triplets. For each triplet, please provide the following details: **Aspect:** The specific attribute, feature, or component mentioned in the text (e.g., battery life, customer service). **Opinion:** A descriptive phrase or clause that expresses the evaluation or opinion about the aspect in the text. **Sentiment Polarity:** The overall sentiment orientation as inferred from the semantic context (positive, negative, neutral). Input sentence: $text **Requirements:** Ensure that each triplet accurately reflects the semantic meaning of the text. Do not omit any aspect or opinion that is clearly expressed.	Please analyze the following text from the perspective of explicit sentiment expressions. Your task is to extract aspect-based sentiment triplets by focusing on sentiment words and phrases. For each triplet, include the following details: **Aspect:** The specific aspect or feature mentioned (e.g., software stability, customer support). **Opinion:** The key word or phrase that directly expresses sentiment (e.g., "excellent," "disappointing"). **Sentiment Polarity:** A detailed description of the sentiment, clearly classifying it as positive, negative, or neutral, based on the sentiment word and context. Input sentence: $text **Requirements:** Focus on explicit sentiment cues in the text. Make sure to link the sentiment word directly to the relevant aspect.

Example:
Input: The screen of this phone is very clear, but the battery life is a bit short.
Output: [screen, clear, positive], [battery, short, negative]

Figure 3. The semantic and sentiment prompt template.

Using the semantic prompt template, the LLM can generate aspect-based triplets that capture the overall meaning of a sentence. For instance, given the input sentence "this vase is very beautiful", the LLM produces the output:

$$T_{LLM(sem)}^{text} = [vast, beautiful, positive] \tag{5}$$

3.3.2. Sentiment Prompt Template

This template focuses on explicit sentiment expressions within a sentence. It instructs the model to precisely identify sentiment-bearing words and to associate these words with their corresponding aspects while also providing a sentiment classification (i.e., positive, negative, or neutral). As illustrated in Figure 3 for the sentiment prompt template, when this prompt is fed into the LLM, it produces structured aspect-based sentiment triplets. The output triplets generated by the LLM based on the sentiment prompt template can be represented as follows:

$$T_{LLM(sen)}^{text} = LLM(X_{prompt}^{sen}) \tag{6}$$

where $LLM(\cdot)$ represents large language model, and X_{prompt}^{sen} represents sentiment prompt template.

In our extraction process, we concurrently employ both prompt templates to obtain triplet outputs from semantic and sentiment perspectives. We then fuse these outputs by taking their intersection, thereby mitigating potential biases and extraction errors that may arise from relying on a single viewpoint. This can be represented as follows:

$$T_{LLM}^{text} = T_{LLM(sem)}^{text} \cap T_{LLM(sen)}^{text} \tag{7}$$

This dual-view prompt template extraction method leverages the potent language understanding and generative capabilities of LLMs while mitigating the error propagation issues inherent in traditional single-strategy approaches by integrating and fusing information from multiple perspectives. Consequently, it offers an efficient and robust solution for aspect-based sentiment extraction task.

3.3.3. Triple Enhanced Context Encoding

The aspect-based sentiment triplets are fed into BERT to generate semantic feature vector representations, thereby enhancing the contextual representation of the sentence in the subsequent step. The formulation is as follows:

$$H_{T_{LLM}} = BERT(\{T_{LLM}^{text}\}_{i=1}^{N}) \tag{8}$$

where N represents the number of aspect-based sentiment triple. The semantic representations of T_{LLM}^{text} are encoded as vector embeddings and passed to the context enhancement module to enrich contextual information.

To enhance the capability of model in identifying aspect terms, we integrate part-of-speech (POS) into word embeddings, recognizing that aspect terms are predominantly nouns [27]. Specifically, we first employ natural language toolkit NLTK [28] to annotate each word in the sentence with its corresponding POS tag. Subsequently, we construct the token representation by summing the token embedding, positional embedding, and POS tag embedding. This integration not only enriches the input representation but also improves the precision in recognizing aspect terms.

$$H = BERT(\langle CLS \rangle W \langle SEP \rangle \oplus POS - t \oplus posi - em) \tag{9}$$

where $POS - t$ represents POS information, and posi-em represents positional information. $H \in \mathbb{R}^{n \times d}$, where n is the number of tokens. The multi-head attention mechanism is employed to integrate $H_{T_{LLM}}$ into H, generating the enhanced contextual representation H_{aug}. $\oplus$ represents the element-wise addition of token embeddings, POS tag embeddings, and positional embeddings to form the final input representation fed into BERT.

$$H_{aug} = H + MultiHeadAttention(Q = H, K = H_{T_{LLM}}, V = H_{T_{LLM}}) \tag{10}$$

where $MultiHeadAttention(\cdot)$ represents the multi-head attention mechanism introduced by Vaswani et al. [29]. The loss function for context enhancement based on large-language-model-extracted triplets can be defined as follows:

$$L_{LLM} = \sum_{t \in T_{LLM}} - \log P(t|H_{aug}) \tag{11}$$

3.4. Exploiting Duality in Aspect Sentiment Triplet Extraction

Ref. [11] proposed a bidirectional extraction method based on sequential prompting to address the ASTE task, achieving remarkable results. Building on their approach, this study employs a bidirectional strategy, consisting of two directions: first, extracting aspect terms, followed by sentiment polarity and opinion terms (A2SO); second, extracting opinion terms, followed by aspect terms and sentiment polarity (O2AS). By performing triplet extraction in both directions simultaneously, the model can more comprehensively and accurately capture sentiment information in the text, thereby enhancing overall performance.

3.4.1. A2SO
Extract Aspect Terms

To enhance the expressiveness of prompts, Ref. [11] designs an aspect-aware hybrid prompt. This prompt consists of discrete tokens and virtual tokens (trainable continuous embeddings) and can be represented as follows:

$$Prompt_a = [[u_{a,1}], \dots, aspect, \dots, [u_{a,x}]] \tag{12}$$

where u_{a_i} represents a virtual token from the vocabulary of the BERT pre-trained language model, and x represents the number of virtual tokens. The aspect-aware prompt input can be formulated as follows:

$$Input_a = [\langle CLS \rangle, Prompt_a, \langle SEP \rangle] \tag{13}$$

Encode it into feature vectors using BERT:

$$H_a = BERT(Input_a) \tag{14}$$

Concatenate the enhanced context feature representations with the vector representations of the aspect prompts.

$$H_{fin} = H_a \oplus H_{aug} \tag{15}$$

To effectively extract aspect terms from sentences, we employed the Global Pointer Network (GPN) [26]. This method utilizes the GPN to effectively extract aspect terms from sentences.

$$P_{A2SO}(a_{start}, a_{end}) = GPN(H_{fin}) \tag{16}$$

where GPN represents the Global Pointer Network model. a_{start} and a_{end} represent the indices for the start and end of aspect terms, respectively.

Extract Sentiment Polarity and Opinion Terms

Taking inspiration from [26], we transform the problem of sentiment polarity into a mask prediction problem. Specifically, we first construct language prompts based on the extracted aspect terms, as depicted below:

$$Prompt_s(a) = [[u_{s,1}], \ldots, a, \ldots, [MASK], \ldots, [u_{s,y}]] \tag{17}$$

where $[u_s, i]$ represents a virtual token, and a represents aspect term. We apply the BERT model to obtain feature representations H_s based on aspect term prompts. These are then concatenated with the triple-sentence feature representations generated by the large model, resulting in $H_{sas} = H_s \oplus H_{aug}$. According to the following formula, we derive the sentiment polarity of the $[mask]$ feature representation in H_{sas}:

$$P(s|a, W) = softmax(W_s h_{[MASK]} + b_s) \tag{18}$$

where W_s and b_s are trainable parameters. After successfully extracting aspect terms and sentiment polarity, the next step involves extracting opinion words from the sentences. To achieve this, a novel prompting approach designed specifically tailored for extracting opinion words from sentences.

$$Prompt_o(a, s) = [[u_{o,1}], \ldots, a, \ldots, s, \ldots, [u_{o,y}]] \tag{19}$$

where $[u_o, i]$ represents a virtual token, a represents an aspect term, and s represents sentiment polarity. Similar to the method for extracting aspect terms, we concatenate the prompted feature representations with the contextual feature representations derived from the triplets of the LLM, culminating in the final feature representation. Subsequently, we employ the GPN model to ascertain the start and end indices of opinion words, thereby finalizing the extraction process. This meticulous approach ensures that opinion word extraction fully harnesses existing context and prompt information, thereby enhancing accuracy and robustness. In the A2SO extraction stage, the GPN model is employed, and the loss function is defined as follows:

$$Loss_{A2SO} = -(\sum_{t \in T_{true}} \log P_{A2SO}(t.a) + \log P_{A2SO}(t.s|t.a) + \log P_{A2SO}(t.o|t.a, t.s)) \quad (20)$$

where T_{true} represents the label of triplet.

3.4.2. O2AS

The O2AS method first extracts opinion words and then identifies the corresponding aspect terms and sentiment polarity. The primary difference between O2AS and A2SO lies in the order of the prompting strategy. Therefore, this section explores the design of prompts specifically for the O2AS method.

The prompt for extracting opinion words from a sentence is defined as follows:

$$Prompt_o = [[v_{o,1}], \ldots, opinion, \ldots, [v_{o,y'}]] \quad (21)$$

where $[v_o, 1]$ represents a virtual token. The extraction process follows the same steps as the A2SO method. After successfully extracting opinion words, we further design new prompts to extract the corresponding aspect terms from the sentence.

$$Prompt_a(o) = [[v_{a,1}], \ldots, o, \ldots, [v_{a,y'}]] \quad (22)$$

Finally, after successfully extracting opinion words and aspect terms, the sentiment polarity associated with the aspect terms is further predicted. To achieve this, we carefully design sentiment polarity-specific prompts, which are formulated as follows:

$$Prompt_s(a, o) = [[v_{a,1}], \ldots, o, \ldots, s, \ldots, [mask] \ldots, [v_{a,y'}]] \quad (23)$$

In the O2AS method, the GPN model is applied to predict aspect terms, with the corresponding loss function defined as follows:

$$loss_{O2AS} = -(\sum_{t \in T_{true}} \log P_{O2AS}(t.o) + \log P_{O2AS}(t.a|t.o) + \log P_{O2AS}(t.s|t.o, t.a)) \quad (24)$$

After extracting triples in both the A2SO and O2AS directions, we integrate the results using an intersection operation to ensure consistency and accuracy. The final set of triplets $T = T_{A2SO} \cap T_{O2AS}$, where T_{A2SO} and T_{O2AS} represent the sets of extracted triplets, comprising aspect terms, opinion words, and sentiment polarity, produced by the A2SO and O2AS methods, respectively. This intersection mechanism effectively consolidates information from both extraction directions, enhancing the reliability of the final triplet set.

3.5. Multiple Iterations

The bidirectional triplet extraction model follows a sequential extraction strategy, where dependencies on previous outputs (e.g., extracting aspects first, followed by sentiment and then opinions) may lead to cascading errors. To mitigate this issue, we introduce a multi-round iterative algorithm, allowing the model to refine its predictions across multiple iterations. This algorithm iteratively updates the context and triplet set based on confidence scores, aiming to correct errors and enhance result consistency. Each iteration involves extracting triplets, evaluating their reliability, updating the context with high-confidence results, and re-extracting low-confidence predictions.

The goal of the algorithm is to accurately extract aspect–opinion–sentiment triplets from a given sentence, such as "service-excellent-positive". By iteratively refining the results, the final output consists of reliable triplets. The detailed steps are as follows:

1. Initialization: A LLM generates initial triplet predictions based on two-view prompts and the input sentence. These preliminary predictions, combined with fundamental sentence information, form an enhanced initial context.

2. Multi-Round Iteration: In each iteration, triplets are extracted using two complementary strategies: one follows the aspect-first approach, progressively identifying sentiment and opinion (A2SO), while the other starts with opinions and reversely identifies aspects and sentiment (O2AS). At the end of each iteration, confidence scores are assigned to the extracted triplets, and results from both strategies are merged.

3. High-confidence triplets are selected to update the base information, and the next iteration begins.

The details of the multi-round iteration process are presented in Algorithm 1. Here, H represents the feature representation of the sentence obtained through BERT, and $H_{aug}^{(0)}$ represents the prior feature representation extracted by the LLM.

Algorithm 1 Multiple iterations

Input: H, $H_{aug}^{(0)}$, Maximum number of iterations K, High confidence threshold θ
Output: $T_{final} = (a, o, s)$

1: Initialize $T_{final} = []$, $cache = \{\}$
2: **for** i in range $(0, K)$ **do**
3:     # Extract the triplet using A2SO and O2AS pipeline, and fuse the results.
4:     $T_{A2SO}^{(i)} = extract_{A2SO}(H \oplus H_{aug}^{(i)})$
5:     $T_{O2AS}^{(i)} = extract_{O2AS}(H \oplus H_{aug}^{(i)})$
6:     **for** t in $T_{A2SO}^{(i)} \cap T_{O2AS}^{(i)}$ **do**
7:         $S^{(i)} = \frac{1}{2} * P_{A2SO}^{(i)} + \frac{1}{2} * P_{O2AS}^{(i)}$
8:         $T^{(i)}.append(t)$
9:     **end for**
10:     # Separate high and low confidence triplets
11:     $T_{high}^{(i)} = \{t \in T^{(i)} | s \geq \theta\}$
12:     $T_{low}^{(i)} = \{t \in T^{(i)} | s < \theta\}$
13:     # Update and enhance context
14:     **for** t in $T_{high}^{(i)}$ **do**
15:         $cache[t] = BERT(t)$
16:     **end for**
17:     $H_{aug}^{(i+1)} = H + MultiHeadAttenrion(H, cache.values(), cache.values())$
18:     $T^{(i+1)} = T_{low}^{(i)}$
19: **end for**
20: $T_{final} = cache.keys() \cup T_{low}^{(K)}$
21: Return T_{final}

3.6. Model Training

To integrate the LLM, A2SO, and O2AS methods, a joint training strategy is adopted. Specifically, a comprehensive loss function is constructed by summing the loss of the sentence semantic representation generated by the LLM, the losses from extracting various elements in both the A2SO and O2AS directions, and a regularization term. By minimizing this loss function, the model simultaneously optimizes the semantic understanding capability of the LLM and the extraction accuracy of elements in both directions, thereby achieving effective integration and collaborative training among the three components, ultimately enhancing overall performance.

$$L = \sum_{k=0}^{K} (L_{A2SO}^{k} + L_{O2AS}^{k}) + \lambda \sum_{k=0}^{K} L_{reg}^{k} + \alpha L_{LLM} \qquad (25)$$

where coefficients λ and α are hyperparameters that balance the KL-divergence regularization term and the LLM-enhanced context loss against the main directional extraction losses.

4. Experiments

In this section, we present a comprehensive evaluation of the performance improvements achieved by our proposed framework for ASTE across multiple datasets. To demonstrate the efficacy of our approach, we first provide an extensive overview of the datasets used and detail our experimental settings. Building on this foundation, we then conduct an in-depth analysis of the results to elucidate the key factors driving the observed performance enhancements.

4.1. Datasets

To evaluate the performance of our proposed model, we employ three widely used datasets: ASTE-DATA-V1 [5], ASTE-DATA-V2 [6], and TOWE [30].

ASTE-DATA-V1: The dataset was refined by Peng et al. [5], based on the datasets originally proposed in the SemEval Challenges by Pontiki et al. [31], with opinion annotations sourced from Fan et al. [30]. It comprises four subsets: 14 res, 14 lap, 15 res, and 16 res. The dataset covers real-world reviews in the domains of laptops and restaurants. Each sample includes the original sentence, a sequence labeled with unified aspect tags, and a sequence labeled with opinion tags. A single sentence may contain multiple aspects and corresponding opinions.

ASTE-DATA-V2: The dataset employed in this study builds upon the ASTE-DATA-V1 and was further refined by Xu et al. [6]. It comprises four benchmark subsets: 14 res, 14 lap, 15 res, and 16 res. In contrast to the original ASTE-DATA-V1, the revised version explicitly accommodates cases where a single opinion span is linked to multiple aspect terms. This enhancement better captures the intricacies of real-world sentiment expression and the many-to-one relationships commonly observed between opinions and targets in natural language.

TOWE: The datasets are derived from the SemEval Challenge series Task 4 of SemEval 2014, Task 12 of SemEval 2015, and Task 5 of SemEval 2016 by Pontiki et al. [31]. These benchmark datasets span the restaurant and laptop domains and are widely adopted in various sub-tasks of ABSA, including aspect category detection, opinion term extraction, and opinion-dependent sentiment classification.

These datasets, spanning diverse domains, provide a comprehensive benchmark for assessing the generalization ability and accuracy of the model in ASTE. Each dataset comprises a rich collection of textual samples annotated with aspect terms, opinion words, and sentiment polarity, thereby serving as a reliable standard for evaluating practical performance. By testing our approach across these datasets, we gain a deeper understanding of the strengths and weaknesses of the model, which in turn guides further optimization and refinement. Table 1 details the basic characteristics of these datasets, including the sizes of the training and test sets and the number of aspect terms, among other key metrics. No.sen represents the number of sentences and No.Asp represents the number of aspects.

Table 1. The datasets used to evaluate our proposed model.

Dataset		14 res		14 lap		15 res		16 res	
		No.sen	No.Asp	No.sen	No.Asp	No.sen	No.Asp	No.sen	No.Asp
ASTE-V1	Train	1300	2145	920	1265	593	923	842	1289
	Dev	323	524	228	337	148	238	210	316
	Test	496	862	339	490	318	455	320	465
ASTE-V2	Train	1266	2338	906	1460	605	1013	857	1394
	Dev	310	577	219	346	148	249	210	339
	Test	492	994	328	543	322	485	326	514
TOWE	Train	1625	3066	1150	1881	754	1277	1079	1770
	Test	500	1030	343	567	325	493	328	525

4.2. Experiment Setting

During training, we employ the AdamW optimizer with a learning rate of 5×10^{-5} and train the model for 20 epochs [32]. A warm-up strategy is applied during the first 10% of training to accelerate convergence and enhance stability [29], and a batch size of 12 is used to balance training efficiency with comprehensive feature extraction. To ensure the robustness and statistical reliability of our experimental results, we repeat each experiment five times using different random seeds and report the mean values. In the LLM generation phase, the DeepSeek-R1 (DeepSeek: https://api.deepseek.com, accessed on 25 January 2025) model is utilized to generate aspect sentiment triplets. DeepSeek-R1 was selected as the backbone LLM for this study due to its strong performance, open-source availability, and cost-effectiveness. For the multi-round iterative process, the maximum number of iterations K is set to 3, and the confidence threshold is fixed at 0.80. The coefficients $\lambda = \alpha = 1$.

In line with [9,11,33], our evaluation framework encompasses not only the triplet extraction task but also three additional tasks to thoroughly assess model performance. To evaluate the effectiveness of our proposed model and ensure a fair comparison with state-of-the-art methods, we adopt three widely recognized evaluation metrics: precision (P), recall (R), and F1-score (F1). These metrics are particularly suited for the ASTE task, where correct identification of all three elements, aspect term, opinion term, and sentiment polarity, is required. Precision reflects the ability of model to produce only correct triplets, thereby quantifying its strictness in prediction. Recall measures the capacity of model to retrieve all relevant triplets, highlighting its coverage. F1-score, as the harmonic mean of precision and recall, offers a balanced view and is especially important when dealing with imbalanced data distributions common in real-world ABSA datasets.

A triplet is considered correct only if the aspect term, opinion term, and sentiment polarity all match the ground truth exactly, making these metrics both rigorous and appropriate for evaluating fine-grained sentiment extraction performance. Following previous works [11,15,34], we report these metrics to maintain consistency and comparability across benchmarks. All experimental evaluations are executed on a GPU (Geforce RTX 3090, 24 GB).

4.3. Baselines

While an exhaustive comparison with every existing ASTE method exceeds the scope of this paper, we have selected representative and widely cited models from two major categories:

For two-stage extraction methods in aspect-based sentiment triplet extraction, baselines include CMLA+, RINANTE+, and Li-unified-R [5], which are enhanced variants of CMLA [35], RINANTE [36], and Li-unified [37], respectively. In addition, the methods

proposed by Peng-TG [5] and TGCN [38] are also notable. The Peng-TG model first extracts aspects and opinions, subsequently pairing them, with sentiment being identified in the initial stage; in contrast, TGCN predicts sentiment based on the paired outputs. DAST [11] introduces two distinct directional triplet extraction strategies that derive triplets from different combinations. These approaches often suffer from cascading errors due to their sequential design, where inaccuracies in one stage propagate to the next.

Regarding joint extraction approaches, OTE-MTL [17] and Span ASTE [13] employ multi-task learning to jointly extract all three elements of ASTE. Moreover, specialized tagging frameworks for triplet prediction have been proposed by GTS [7], JET [6], and Jing-SC [8]. In addition, Dual-MRC [9] and Bi-MRC [10] leverage MRC-based designs for triplet prediction, while generative frameworks with pointer indexing have been developed by Yan-UG [14] and SentiPrompt [4]. Furthermore, GAS [15] and PARAPHRAS [34] model the ASTE task as a text generation problem, thereby generating all relevant elements from a sentence. While these avoid cascading errors, they tend to struggle with long-range dependencies and complex multi-sentiment targets.

These methods reflect the current state-of-the-art and predominant approaches used in the literature and demonstrate distinct advantages and performance characteristics across diverse application scenarios and datasets, providing a rich array of solutions for the ASTE task. Our comprehensive comparative analysis against these representative baselines effectively demonstrates the robustness and superior performance of our proposed DASTER framework.

4.4. Experimental Results

In this section, we present a detailed analysis of the experimental results on the AV2 dataset, the results shown in Table 2. Our proposed model is compared against multiple state-of-the-art methods across four datasets. We employ precision (P), recall (R), and F1-score (F1) as evaluation metrics to ensure a comprehensive and accurate assessment of the effectiveness of the model.

Table 2. The results of experiment on AV2.

Methods	14 res			14 lab			15 res			16 res		
	P	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
CMLA+ [5]	39.18	47.13	42.79	30.09	36.92	33.16	34.56	39.84	37.01	41.34	42.10	41.72
RINANTE+ [5]	31.42	39.38	34.95	21.71	18.66	20.07	29.88	30.06	29.97	25.68	22.30	23.87
Li-unified-R [5]	41.04	67.35	51.00	40.56	44.28	42.34	44.72	51.39	47.82	37.33	54.51	44.31
Peng-TG [5]	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	46.96	64.24	54.21
TGCN [38]	63.59	73.44	68.16	57.84	59.33	58.58	54.53	63.30	58.59	63.57	71.98	67.53
OTE-MTL [17]	63.07	58.25	60.56	54.26	41.07	46.75	60.88	42.68	50.18	65.65	54.28	59.42
Span-ASTE [13]	72.89	70.89	71.85	63.44	55.84	59.38	62.18	64.45	63.27	69.45	71.17	70.26
GTS [7]	68.71	67.67	68.17	58.54	50.65	54.30	60.69	60.54	60.61	67.39	66.73	67.06
JET-T [6]	63.44	54.12	58.41	53.53	43.28	47.86	68.20	42.89	52.66	65.28	51.95	57.85
JET-O [6]	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
Jing-SC [8]	67.95	71.23	69.55	62.12	56.38	59.11	58.55	60.00	59.27	70.65	70.23	70.44
Dual-MRC [9]	70.62	69.64	70.28	58.52	55.68	57.06	62.52	56.76	59.50	67.10	67.96	67.52
Bi-MRC [10]	70.26	69.28	69.76	62.46	55.28	57.67	62.80	59.55	60.86	68.42	69.64	69.03
Yan-UG [14]	65.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
SentiPrompt [4]	72.79	72.94	72.86	63.40	58.60	60.90	62.97	62.06	62.51	70.20	73.35	71.74
GAS [15]	72.81	71.56	72.18	64.52	57.27	60.68	63.36	60.62	61.96	69.26	71.15	70.19
PARAPHRASE [34]	71.61	72.41	72.01	63.44	59.12	61.20	60.70	64.33	62.46	70.34	73.35	71.81
DAST [11]	76.75	71.14	73.84	69.26	55.82	61.82	69.92	60.39	64.81	75.60	72.54	74.04
ours	79.32	73.57	76.29	71.61	58.33	63.46	71.87	63.28	67.01	76.34	75.25	76.40

The experimental results demonstrate the outstanding performance of our proposed method across all datasets, achieving the highest F1-scores on each dataset. Specifically, on the 14res dataset, the F1-score reaches 76.29%; on the 14 lab dataset, it is 63.46%; on the 15 res dataset, it attains 67.01%; and on the 16 res dataset, it even reaches 76.40%. These results consistently surpass all baseline models, underscoring the efficiency of our method in the ASTE task. In addition, to rigorously assess the significance of performance gains of our approach over DAST, we conduct two-tailed t-test on the 14 res, 14 lab, 15 res, and 16 res datasets. The resulting p-values, 0.008, 0.040, 0.019, and 0.037 ($p < 0.5$), demonstrating that the improvements in F1-score are statistically significant. These findings substantiate the robustness of our approach relative to DAST.

For the 14 res dataset, our model achieves a precision of 79.32%, a recall of 73.57%, and an F1-score of 76.29%, which are 2.45% and 4.44% higher than those of DAST and Span ASTE, respectively. On the 14lab dataset, although our model does not attain the highest recall, it achieves the best precision and F1-score. In the 15res dataset, the recall of our model is slightly lower than that of the Span-ASTE model by 1.17%, yet it significantly outperforms Span-ASTE in terms of precision and F1-score, with improvements of 9.69% and 3.74%, respectively. In the 16res dataset, the performance gap is particularly pronounced; our method consistently outperforms baselines, achieving an F1-score of 76.40%, which is 2.36% higher than that of DAST, which achieved 74.04%.

Notably, recent methods such as DAST, SentiPrompt, PARAPHRASE, and GAS have demonstrated strong performance on the AV2 dataset; however, our method consistently outperforms them across all metrics. For example, compared to ASTE, our model achieves a 2.45% improvement in the F1-score on the 14res dataset, indicating that our framework effectively captures contextual dependencies and structured information to enable more accurate sentiment triplet extraction. Moreover, although prompt-based approaches such as SentiPrompt and PARAPHRASE exhibit commendable performance, they remain inferior to our model. Specifically, on the 16res dataset, our model improves the F1-score by 4.66% and 4.59% compared to SentiPrompt and PARAPHRASE, respectively. These results provide strong evidence for the effectiveness of integrating LLM with bidirectional methods, which better model syntactic structure and capture fine-grained aspect–opinion relationships.

Further analysis of precision, recall, and F1-score reveals that our model achieves a well balanced between high recall and high precision. In contrast, several baseline models exhibit a pronounced trade-off between these metrics. For instance, the Li-unified-R model and Peng-TG obtain high recall at the cost of low precision, resulting in modest F1-scores, whereas the JET-T and JET-O models favor precision at the expense of recall, which undermines overall performance. We attribute the superior performance of our proposed model to two key factors. First, our approach leverages a LLM as prior knowledge within a bidirectional extraction framework, effectively mitigating cascading errors. Second, a confidence-based iterative optimization strategy not only enhances classification accuracy but also incorporates high-confidence samples into training, thereby compensating for the limitations inherent in two-stage methods for triplet extraction.

4.5. Ablation Study

To evaluate the contribution of each component in the model, we perform an ablation study. Specifically, we systematically remove individual components to assess the combined effect of the remaining elements. The analysis results are presented in Table 3.

Table 3. The F1-score of ablation study.

Method	14 res	14 lap	15 res	16 res
w/o LLM	75.23	63.01	66.34	75.73
w/o Iterration	74.91	62.75	65.89	75.31
w/o LLM and Iteration	73.84	61.82	64.81	74.04
DASTER	76.29	63.46	67.01	76.40

The ablation study in Table 3 evaluates the individual contributions of the LLM and iterative components within the DASTER framework. When the LLM component is removed, F1-scores across all subsets decline; for example, the F1-score on the 14res subset drops from 76.29% to 75.23%, indicating that the LLM is crucial for capturing subtle linguistic features. Similarly, removal of the iterative module results in a decrease in F1-scores, with the 14res score falling to 74.91%. Notably, simultaneous removal of both components leads to a substantial performance drop (e.g., the 14res score decreases to 73.84%), highlighting the complementary nature of these modules. In contrast, the DASTER model achieves the highest F1-scores across all datasets, demonstrating that integration of the LLM and iterative reinforcement is essential for effectively capturing syntactic and semantic information. These results provide strong evidence for the robustness and efficacy of the proposed method in enhancing ASTE extraction.

5. Analysis and Discussion

In this section, we not only conduct experiments on additional subtasks but also perform parameter analysis, case studies and error analysis to further validate the effectiveness of our model.

5.1. Effect of the Parameters

In this subsection, we investigate the sensitivity of the iteration count K through experiments and analyze how different hyperparameter settings affect our results. Specifically, we carried out two key studies: for the effect of iteration count K on the F1-score, we evaluated the performance of the model across various values of K to identify the optimal number of iterations. The results are presented in Figure 4. For the F1-score under different hyperparameter configurations, we tested a range of hyperparameter combinations and examined their impact on model performance, with the aim of determining the best-performing configuration. These findings are shown in Figure 5. These experiments are designed to provide a deep understanding of how hyperparameter selection influences the effectiveness of the models.

Figure 4 illustrates the impact of varying the iteration count k (ranging from 1 to 5) on the F1-scores across different datasets, providing insights into the optimal value of k, convergence behavior, and overall performance trends. Notably, $k = 3$ consistently yields peak or near-peak F1-scores across all datasets, suggesting that a moderate number of iterations effectively enhances the aspect sentiment triplet extraction performance. Furthermore, F1-scores exhibit stability for $k \geq 3$ across all datasets, confirming the numerical stability and convergence of the DASTER framework. This behavior aligns with the theoretical expectation of iterative algorithms, indicating that the model essentially converges by the third iteration and additional iterations provide only marginal gains. We also observe diminishing returns for $k > 3$ (e.g., a slight decrease in the 14res dataset), indicating a saturation effect. This implies that excessive iterations do not lead to further accuracy improvements but instead may increase computational costs without practical benefit.
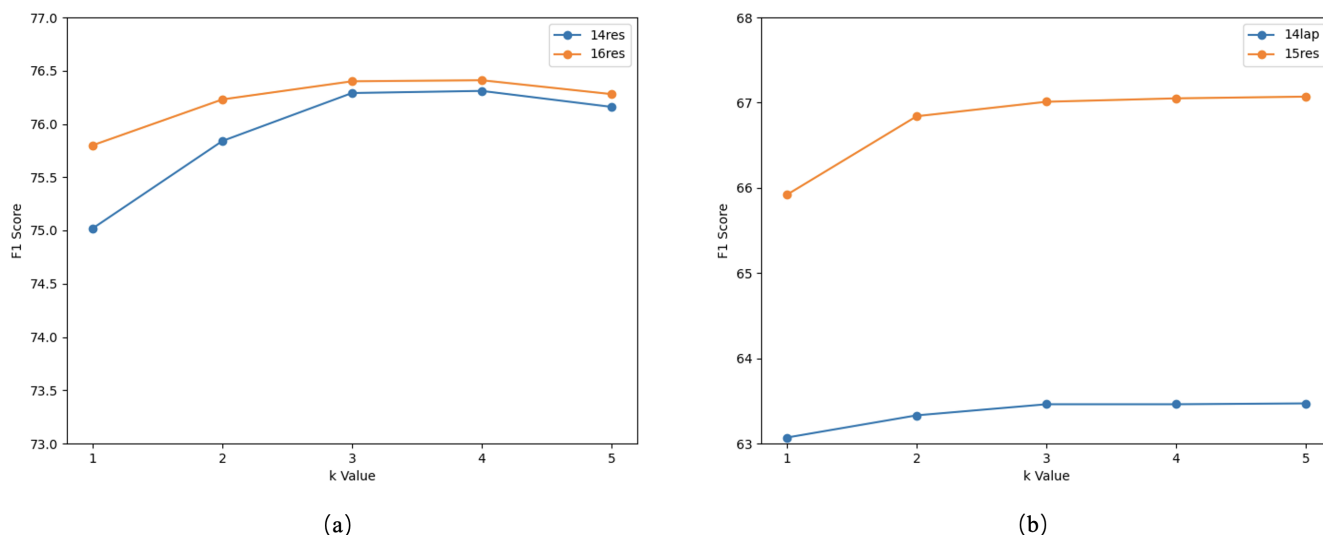
(a)   (b)

Figure 4. Sensitivity of F1-score to iteration count *K* across datasets. (**a**) is the results on 14 res and 16 res. (**b**) is the results on 15 lap and 15 res.
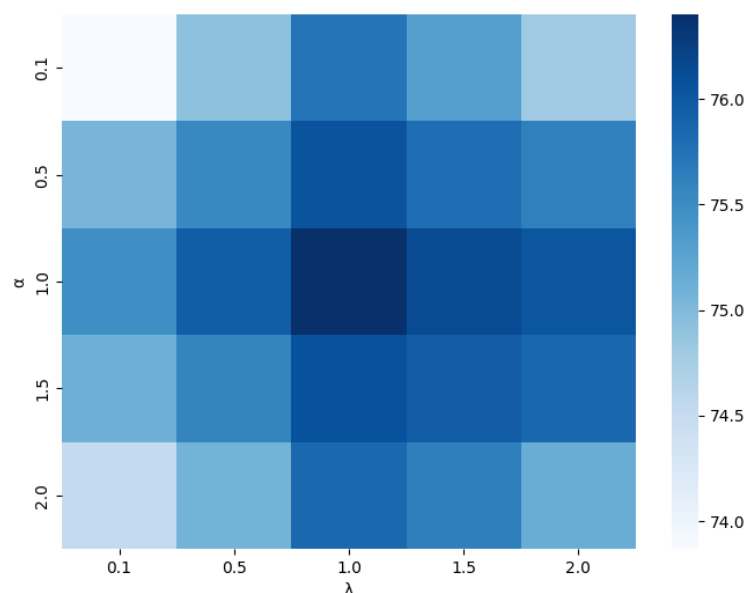

Figure 5. F1-scores on the 16res dataset for different α and λ.

Figure 5 depicts the joint effect of α and λ on the ability of the model to integrate LLM priors and to enforce consistency between the two extraction directions. The heatmap presents the F1-score for α and λ values in the range [0.1, 2.0]. Notable findings are as follows: Lowest performance at ($\alpha = 0.1$, $\lambda = 0.1$), insufficient weighting of the LLM prior combined with weak KL-divergence regularization limits both prior integration and bidirectional output alignment, resulting in the lowest F1-score. When α is set to 0.1, incorporation of the LLM prior is inadequate, thereby constraining the performance of the model and α exceeds 1.5, and the F1-score declines, indicating that excessive reliance on the LLM prior impairs model performance. A regularization coefficient of $\lambda = 0.1$ produces a low F1-score, reflecting divergence between A2SO and O2AS outputs due to inadequate constraint. In contrast, $\lambda = 1.0$ maximizes the F1 by aligning bidirectional predictions without imposing excessive restrictions. For λ greater than 1.5, performance again declines, suggesting that overly strong regularization suppresses beneficial diversity in the extracted triplets. The optimal region at $\alpha = 1.0$, $\lambda = 1.0$, the darkest cell in the

heatmap indicates that a balanced combination of prior weighting and regularization maximizes extraction accuracy.

Overall, these results identify $k = 3, \alpha = 1.0$, and $\lambda = 1.0$ as the optimal hyperparameter configuration for DASTER. This setting achieves an effective balance among iterative refinement, LLM prior integration, and KL-divergence regularization, thereby substantially enhancing both robustness and accuracy in the ASTE task.

5.2. The Results of Subtasks

We evaluated our model against compared models on the AESC and AOE tasks using the AV1 dataset, as shown in Tables 4 and 5, and assessed the results of the EA, EO, and AOE subtasks on the TOWE dataset, as shown in Table 6.

Table 4. The results for AESC task on AV1 dataset.

Methods	14 res			14 lab			15 res			16 res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+	67.80	73.69	70.62	54.70	59.20	56.90	49.90	58.00	53.60	58.90	63.60	61.20
RINANTE+	48.97	47.36	48.15	41.20	33.20	36.70	46.20	37.40	41.30	49.40	36.70	42.10
Li-unified-R	73.15	74.44	73.79	66.28	60.71	63.38	64.95	64.95	64.95	66.33	74.55	70.20
Peng-TG	74.41	73.97	74.19	63.15	61.55	62.34	67.65	64.02	65.79	71.18	72.30	71.73
Dual-MRC	76.84	76.31	76.57	67.45	61.96	64.59	66.84	63.52	65.14	69.18	72.59	70.84
Bi-MRC	77.14	75.10	76.39	72.73	62.59	67.27	72.41	62.63	67.16	73.69	72.69	73.18
DAST	80.56	75.14	77.76	75.03	62.22	68.03	77.74	62.56	69.33	79.83	77.51	78.51
ours	81.82	75.64	78.89	76.37	62.91	69.58	78.70	63.24	71.16	81.02	78.46	79.85

Table 5. The results for AOE task on AV1 dataset.

Methods	14 res			14 lab			15 res			16 res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
CMLA+	45.17	53.42	48.95	42.10	46.30	44.10	42.70	46.70	44.60	52.50	47.90	50.00
RINANTE+	42.32	51.08	46.29	34.40	26.20	29.70	37.10	33.90	35.40	35.70	27.00	30.70
Li-unified-R	44.37	73.67	55.34	52.29	52.94	52.56	52.75	61.75	56.85	46.11	64.54	53.75
Peng-TG	47.76	68.10	56.10	50.00	58.47	53.85	49.22	65.70	56.23	52.35	70.50	60.04
Dual-MRC	76.23	73.67	74.93	65.43	61.43	63.37	72.43	58.90	64.97	77.06	74.41	75.71
Bi-MRC	76.91	75.59	76.23	74.11	61.92	67.45	71.59	65.89	68.60	76.08	76.99	76.52
DAST	80.76	77.71	79.21	77.74	64.52	70.52	77.79	68.21	72.96	81.10	81.70	81.40
ours	81.33	78.69	80.43	79.31	65.40	71.74	79.03	68.94	74.62	82.07	82.37	82.75

Our approach achieves superior F1-scores. For instance, on the 14res dataset, our method attains an F1-score of 78.89%, surpassing the second-best model DAST, which achieves 77.76%. When evaluating bidirectional approaches, such as Bi-MRC A2O and O2A, as well as DAST A2SO and O2AS, and our model, which leverages both A2SO and O2AS, consistently outperforms methods that employ only a single extraction direction (dual-MRC A2O) on all three evaluation metrics across. This improvement can be attributed to the mutual reinforcement between the different directional strategies during joint training. Compared to DAST, our method demonstrates superior performance on nearly all metrics. The performance gains stem from two key factors. First, we leverage the prior knowledge obtained from a LLM with two-view prompt template and integrate it with sentence representations via a multi-head attention mechanism, providing robust prior information for subsequent extraction. Second, we implement a multi-round iterative strategy that selectively incorporates high-confidence samples into subsequent training rounds, thereby

enhancing extraction accuracy. Together, these innovations effectively mitigate cascading errors and contribute to the robust performance of our approach in ASTE.

Table 6. The F1-score (%) for AE, OE, and AOE on TOWE dataset.

Methods	14 res			14 lab			15 res			16 res		
	AE	OE	AOE	AE	OE	AOE	AE	OE	AOE	AE	OE	AOE
HAST	82.56	75.10	62.39	79.14	67.50	53.41	79.84	68.45	58.12	81.44	75.71	63.84
JERE	79.79	77.44	66.02	74.61	64.02	52.34	75.00	71.38	59.64	76.08	78.02	67.65
SpanMit	87.42	83.98	75.60	84.51	80.61	68.66	81.76	78.91	64.68	85.62	85.33	71.78
DAST	86.23	87.19	79.46	82.81	80.57	70.12	81.00	80.63	70.86	85.57	87.55	79.59
ours	87.09	88.05	80.83	83.61	81.20	71.38	81.94	81.45	72.62	86.62	88.47	80.77

Table 6 presents the F1-scores for three primary subtasks on the TOWE dataset: aspect extraction (AE), opinion extraction (OE), and the joint task of aspect and opinion extraction (AOE). Our method is compared against several state-of-the-art baselines, including HAST [30], JERE [39], SpanMit [33], and DAST [11], and is evaluated on four subsets: 14res, 14lab, 15res, and 16res. For the AE subtask, our approach demonstrates robust performance. Although SpanMit slightly outperforms our method on the 14res and 14lab subsets (with scores of 87.42% and 84.51%, respectively), our method achieves superior performance on the 15res and 16res subsets, indicating strong robustness across datasets. In the OE subtask, our method consistently yields higher F1-scores than the baselines. Notably, it outperforms HAST and JERE across all four subsets and performs competitively against DAST and SpanMit. For the joint AOE task, our method attains the highest F1-scores across all subsets. Although DAST exhibits strong performance, it consistently falls short by approximately 1–2% on each subset. Overall, the detailed experimental analysis demonstrates that our proposed method not only enhances performance on individual subtasks but also achieves exceptional joint performance. The comprehensive improvements in F1-scores confirm that our model effectively captures the complex interdependencies between aspects and opinions, thereby reinforcing its potential for advancing AOE task.

5.3. Visualization

Figure 6 presents a visualization of the multi-head attention matrix for the sentence "the broccolis were so fresh and tasty". In this visualization, rows and columns correspond to individual tokens in the sentence, and each cell indicates the attention weight from the token in the corresponding row to the token in the corresponding column. The results show that the adjectives "fresh" and "tasty" exhibit notably high attention weights with "broccolis", indicating that the model successfully associates the aspect (broccolis) with its descriptive opinions, an association that is critical for accurate sentiment triplet extraction. Furthermore, the adverb "so" displays high attention weights in relation to "fresh" and "tasty", underscoring its role in modulating sentiment intensity. In contrast, function words such as "the", "were", and "and" consistently exhibit low off-diagonal weights in the typical range, suggesting that these words contribute minimally to semantic and sentiment interactions, which is desirable in sentiment extraction tasks. Overall, the visualization clearly illustrates that the model emphasizes relationships among semantically rich tokens, especially between the aspect "broccolis" and its associated opinion words "fresh" and "tasty". This pattern validates the effectiveness of the multi-head attention mechanism in integrating syntactic and semantic cues, ultimately facilitating more accurate sentiment triplet extraction.
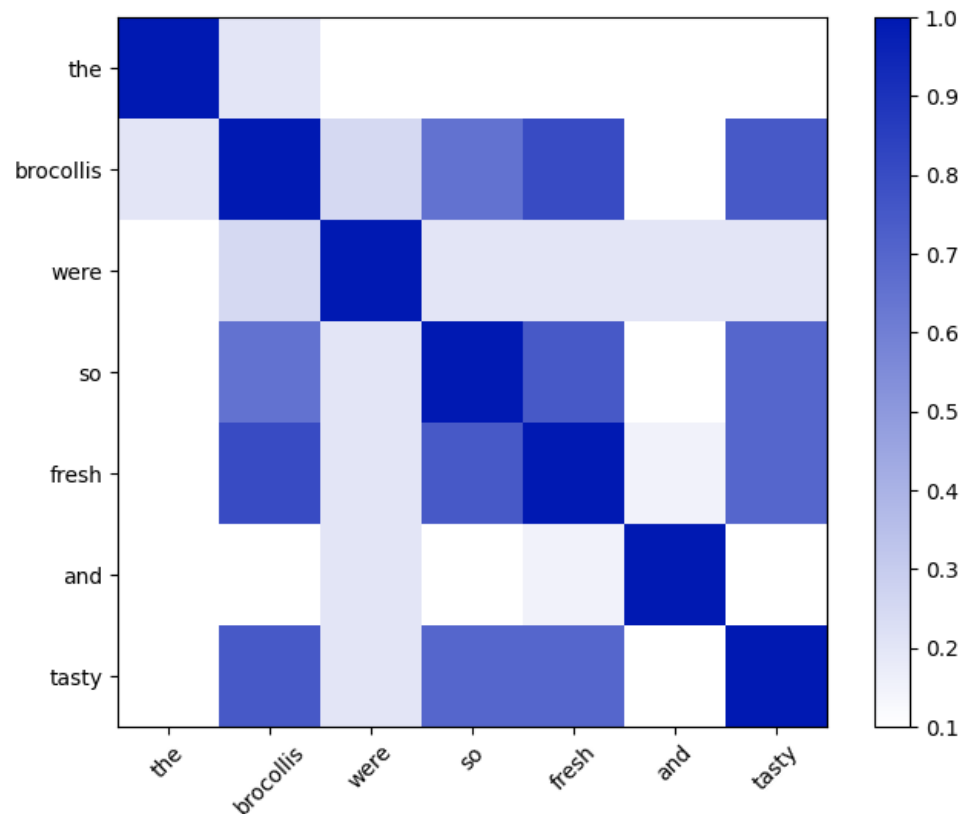
Figure 6. The visualization of the multi-head attention matrix for the sentence "the broccolis were so fresh and tasty".

5.4. Case Study

In the case study, the DAST model was selected as the baseline for comparison. The experimental results are presented in Figure 7.

Id	Sample	results
1	The speed is incredible and I am more than satisfied.	Gold:[speed, satisfied, POS] DAST:[speed, satisfied, POS] ours:[speed, satisfied, POS]
2	Sometimes the screen even goes black on this computer.	Gold:[screen, goes black, NEG] DAST:[screen, goes black, NEG] ours:[screen, goes black, NEG]
3	The service is excellent, the decor is great, and the text color blue food is delicious.	Gold:[service, excellent, POS], [decor, great, POS], [food, delicious, POS] DAST:[service, excellent, POS], [decor, great, POS], [food, delicious, POS] ours: [service, excellent, POS], [decor, great, POS], [food, delicious, POS]
4	The sandwiches are dry tasteless and overprice.	Gold:[sandwiches, dry tasteless overprice , NEG] DAST:[sandwiches, dry tasteless overprice , NEG] ours:[sandwiches, dry tasteless overprice , NEG]
5	Similar to other Indian restaurants, they use the dinner special to attract customers at the door.	Gold: [dinner special, attract customers, NEU] DAST:[dinner special, attract customers, POS] ours:[dinner special, attract customers, NEU]

Figure 7. Case study of proposed model.

Both models successfully predicted the aspect, opinion, and sentiment polarity triplets in the first three sentences, thereby validating the effectiveness of the bidirectional extraction approach. However, in the fourth sentence, the DAST method failed to correctly identify the opinion term, whereas our model not only accurately determined the sentiment polarity but also reliably extracted both the aspect and opinion terms. In the fifth sentence, the triplet (dinner special, attract customers, and NEU) indicates a neutral sentiment; unlike DAST, which mistakenly classified the sentiment as positive, our approach accurately identified it as neutral. These results underscore our model's superior ability to discern subtle emotional cues.

The remarkable performance of our model is attributable to its multi-task shared cascade learning architecture, which effectively integrates two-view prior knowledge from a LLM and enhances prediction accuracy through confidence-based evaluation during iterative cycles. Consequently, our approach excels in the ASTE task.

5.5. Error Analysis

We performed an error analysis of the performance of the DASTER. We chose 30 failure examples for each subsets in AV2 error case, totaling 120, and showed the distribution of the various error types in Figure 8. Through manual inspection, we found that the errors consisted mainly of the following:

- Triplet matching errors: This error category comprises two types. Mismatched triplets, both the aspect and opinion terms are correctly extracted, but their pairing is incorrect. Incomplete triplets, one of the elements (aspect or opinion) is missing. This type is more common because the model only considers an aspect or opinion term successfully extracted when the results from both extraction directions fully agree; if there is any discrepancy between the two directional outputs, the model will not return that element.

- Missing aspect or opinion term: This error type is relatively common; although the extracted triplets are structurally complete, the model fails to recognize all aspect and opinion terms. Such omissions may stem from the model's incomplete comprehension of the input and from limitations in the bidirectional extraction mechanism.

- Aspect or opinion errors: These errors arise when the model erroneously classifies unrelated terms as aspect or opinion tokens. They originate in inadequate semantic comprehension, which causes ordinary nouns or verbs to be misidentified as aspect descriptors in specific contexts. The superficial resemblance of these terms to genuine product or service attributes further compounds the confusion. Moreover, such initial misclassifications precipitate cascading errors, for example, an aspect identification mistake may subsequently lead to erroneous opinion extraction and sentiment polarity prediction.

- Boundary errors: This error indicates that the model fails to precisely locate the target information, causing the generated spans to diverge from the expected boundaries. When predicting aspect terms, opinion terms, or sentiment polarity, an inability to accurately capture their positions gives rise to this error. Such misalignments may result from biases in the model's processing of complex syntactic structures.

- Sentiment polarity errors: The aspect and opinion spans are accurately extracted, and sentiment polarity assignments remain erroneous. Such failures may stem from instances of irony or subtly conveyed sentiment, as well as from terminology that carries domain-specific emotional nuances, which current models find challenging to classification.
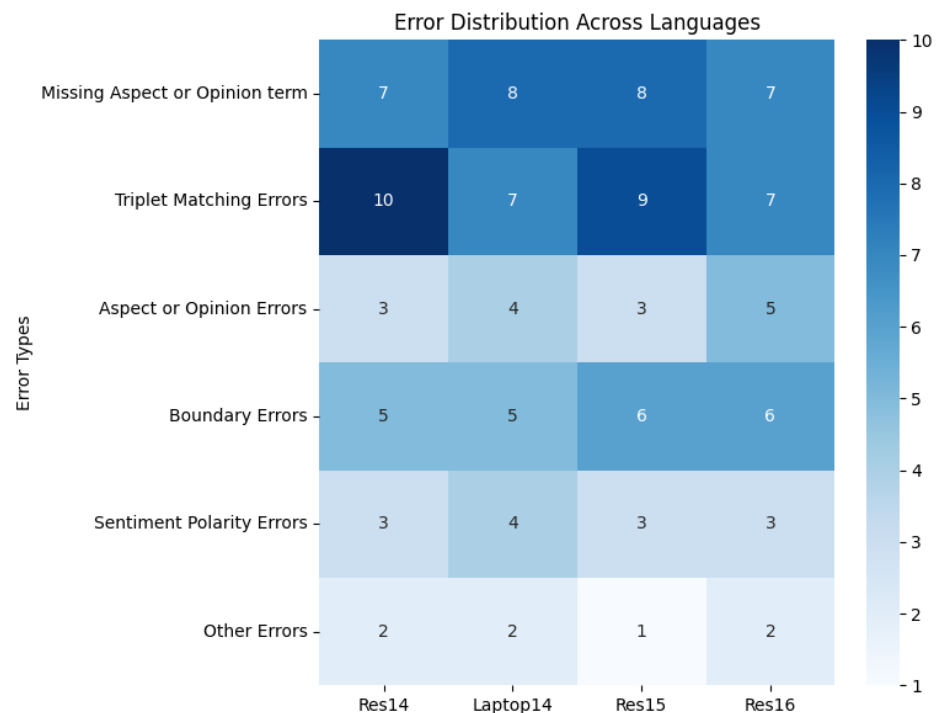
Figure 8. Distribution of errors.

5.6. Practical Implications

The empirical performance of the proposed DASTER framework carries important implications for real-world applications of fine-grained sentiment analysis. In practical settings such as e-commerce platforms, hospitality services, and online review aggregators, users often express opinions involving multiple aspects within a single utterance. Accurately extracting structured sentiment triplets comprising aspect terms, opinion expressions, and sentiment polarity is crucial for understanding nuanced user feedback at scale.

Our method, which integrates dual-directional extraction with iterative optimization guided by confidence thresholds, demonstrates a strong capacity to resolve overlapping and context-dependent sentiment patterns. This is particularly valuable for automatically parsing complex statements such as "The battery lasts long, but the display is disappointing", where multiple contrasting sentiments are associated with distinct aspects. In such cases, the proposed framework preserves semantic precision and improves sentiment alignment.

Moreover, by incorporating LLM reasoning and dynamic reinforcement, the approach exhibits resilience to noise and ambiguity, two prevalent features of user-generated content. These attributes enhance readiness of the model for deployment in multilingual and multi-domain environments, facilitating applications such as customer feedback analysis, product quality monitoring, and real-time opinion tracking. The performance improvements observed on benchmark datasets thus extend beyond theoretical evaluation, offering tangible utility in real-world sentiment-centric systems.

6. Conclusions

In this work, we propose DASTER, a duality-driven framework for aspect-based sentiment triplet extraction that combines LLM with a multi-round iterative refinement process. Our method directly addresses the issue of cascading errors inherent in sequential extraction by leveraging several key mechanisms. First, the dual-view prompt strategy independently extracts candidate triplets from both semantic context and sentiment cues, providing a robust initialization that minimizes significant early errors. Next, the integra-

tion of a multi-head attention mechanism with BERT-encoded sentence features ensures that contextual and sentiment information are thoroughly fused before the sequential extraction begins. Furthermore, a confidence-based iterative optimization process is employed to re-extract low-confidence triplets while preserving high-confidence ones, thereby progressively reducing error propagation. Finally, KL divergence regularization enforces consistency between the two extraction directions, aligning the probability distributions to further mitigate cascading errors.

Extensive experiments on benchmark datasets such as ASTE-DATA-V1, ASTE-DATA-V2, and TOWE demonstrate that DASTER consistently outperforms state-of-the-art approaches. Our results show notable improvements in F1-scores and overall extraction accuracy, confirming the effectiveness of integrating LLM-derived prior knowledge with bidirectional extraction strategies.

Despite these encouraging results, our approach still has certain limitations. Although DASTER performs robustly on standard datasets, its adaptability to low-resource settings, novel domains, and diverse linguistic structures remains underexplored. Moreover, the reliance of framework on sequential pipeline stages introduces potential vulnerabilities to error accumulation.

Future work will focus on developing end-to-end optimization strategies that unify all components of the model within a single training objective and enhancing the scalability and efficiency of the iterative refinement process, particularly the performance of model for large-scale datasets. Additionally, we plan to extend the framework to support domains characterized by implicit sentiment, such as scientific, legal, or social media texts, in order to further improve its generalization capabilities and enhance the robustness of ASTE in real-world scenarios. These efforts aim to extend the applicability of the proposed framework to more diverse and complex real-world scenarios while maintaining robust and precise performance.

Author Contributions: Conceptualization, X.L. and K.Z.; methodology, X.L.; software, X.L.; validation, X.L., K.Z. and D.H.; investigation, K.Z.; data curation, K.Z.; writing—original draft preparation, X.L.; writing—review and editing, X.L., K.Z. and D.H.; visualization, D.H.; supervision, K.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Jiang, Q.; Chen, L.; Xu, R.; Ao, X.; Yang, M. A challenge dataset and effective models for aspect-based sentiment analysis. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 6280–6285.
2. Zarandi, A.K.; Mirzaei, S. A survey of aspect-based sentiment analysis classification with a focus on graph neural network methods. *Multimed. Tools Appl.* **2024**, *83*, 56619–56695.
3. Mughal, N.; Mujtaba, G.; Kumar, A.; Daudpota, S.M. Comparative analysis of deep natural networks and large language models for aspect-based sentiment analysis. *IEEE Access* **2024**, *12*, 60943–60959. [CrossRef]
4. Yin, W.; Liu, C.; Xu, Y.; Wahla, A.R.; Yiting, H.; Zheng, D. SynPrompt: Syntax-aware Enhanced Prompt Engineering for Aspect-based Sentiment Analysis. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), Torino, Italy, 20–25 May 2024; pp. 15469–15479.
5. Peng, H.; Xu, L.; Bing, L.; Huang, F.; Lu, W.; Si, L. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 8600–8607.

6. Xu, L.; Li, H.; Lu, W.; Bing, L. Position-Aware Tagging for Aspect Sentiment Triplet Extraction. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; pp. 2339–2349.

7. Wu, Z.; Ying, C.; Zhao, F.; Fan, Z.; Dai, X.; Xia, R. Grid Tagging Scheme for Aspect-oriented Fine-grained Opinion Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 16–20 November 2020; pp. 2576–2585.

8. Jing, H.; Li, Z.; Zhao, H.; Jiang, S. Seeking Common but Distinguishing Difference, A Joint Aspect-based Sentiment Analysis Model. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 3910–3922.

9. Mao, Y.; Shen, Y.; Yu, C.; Cai, L. A joint training dual-mrc framework for aspect based sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 13543–13551.

10. Chen, S.; Wang, Y.; Liu, J.; Wang, Y. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; Volume 35, pp. 12666–12674.

11. Liu, J.; Chen, T.; Guo, H.; Wang, C.; Jiang, H.; Xiao, Y.; Wu, B. Exploiting duality in aspect sentiment triplet extraction with sequential prompting. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 6111–6123. [CrossRef]

12. Li, X.; Li, D.; Du, R.; Chen, D.; Madden, A. Double policy network for aspect sentiment triplet extraction (student abstract). In Proceedings of the AAAI Conference on Artificial Intelligence, Washington, DC, USA, 7–14 February 2023; Volume 37, pp. 16256–16257.

13. Xu, L.; Chia, Y.K.; Bing, L. Learning Span-Level Interactions for Aspect Sentiment Triplet Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1–6 August 2021; Long Papers; Volume 1, pp. 4755–4766.

14. Yan, H.; Dai, J.; Ji, T.; Qiu, X.; Zhang, Z. A Unified Generative Framework for Aspect-based Sentiment Analysis. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1–6 August 2021; Long Papers; Volume 1, pp. 2416–2429.

15. Zhang, W.; Li, X.; Deng, Y.; Bing, L.; Lam, W. Towards Generative Aspect-Based Sentiment Analysis. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Online, 1–6 August 2021; Short Papers; Volume 2, pp. 504–510.

16. Yuan, L.; Wang, J.; Yu, L.C.; Zhang, X. Encoding syntactic information into transformers for aspect-based sentiment triplet extraction. *IEEE Trans. Affect. Comput.* **2023**, *15*, 722–735. [CrossRef]

17. Zhang, C.; Li, Q.; Song, D.; Wang, B. A Multi-task Learning Framework for Opinion Triplet Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2020, Online, 16–20 November 2020; pp. 819–828.

18. Liu, S.; Li, K.; Li, Z. A robustly optimized BMRC for aspect sentiment triplet extraction. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 10–15 July 2022; pp. 272–278.

19. Chen, H.; Zhai, Z.; Feng, F.; Li, R.; Wang, X. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, Dublin, Ireland, 22–27 May 2022; Long Papers; Volume 1, pp. 2974–2985.

20. Zhang, Y.; Yang, Y.; Li, Y.; Liang, B.; Chen, S.; Dang, Y.; Xu, R. Boundary-driven table-filling for aspect sentiment triplet extraction. In Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing, Abu Dhabi, United Arab Emirates, 7–11 December 2022; pp. 6485–6498.

21. Shi, X.; Hu, M.; Deng, J.; Ren, F.; Shi, P.; Yang, J. Integration of multi-branch GCNs enhancing aspect Sentiment Triplet Extraction. *Appl. Sci.* **2023**, *13*, 4345. [CrossRef]

22. Li, Y.; He, Q.; Zhang, D. Dual graph convolutional networks integrating affective knowledge and position information for aspect sentiment triplet extraction. *Front. Neurorobotics* **2023**, *17*, 1193011. [CrossRef] [PubMed]

23. Peng, K.; Jiang, L.; Peng, H.; Liu, R.; Yu, Z.; Ren, J.; Yu, P.S. Prompt based tri-channel graph convolution neural network for aspect sentiment triplet extraction. In Proceedings of the 2024 SIAM International Conference on data mining (SDM), Houston, TX, USA, 18–20 April 2024; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2024; pp. 145–153.

24. Yang, X.; Peng, T.; Bi, H.; Han, J. Span-level bidirectional retention scheme for aspect sentiment triplet extraction. *Inf. Process. Manag.* **2024**, *61*, 103823. [CrossRef]

25. Zou, W.; Zhang, W.; Wu, W.; Tian, Z. A multi-task shared cascade learning for aspect sentiment triplet extraction using bert-mrc. *Cogn. Comput.* **2024**, *16*, 1554–1571. [CrossRef]

26. Su, J.; Murtadha, A.; Pan, S.; Hou, J.; Sun, J.; Huang, W.; Liu, Y. Global pointer: Novel efficient span-based approach for named entity recognition. *arXiv* **2022**, arXiv:2208.03054.

27. Liu, W.; Lin, S.; Gao, B.; Huang, K.; Liu, W.; Huang, Z.; Huang, F. BERT-POS: Sentiment analysis of MOOC reviews based on BERT with part-of-speech information. In Proceedings of the International Conference on Artificial Intelligence in Education, Durham, UK, 27–31 July 2022; Springer International Publishing: Cham, Switzerland, 2022; pp. 371–374.

28.	Loper, E.; Bird, S. NLTK: The Natural Language Toolkit. In Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics, Philadelphia, PA, USA, 7 July 2002; pp. 63–70.

29.	Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845 aa-Abstract.html (accessed on 12 January 2019).

30.	Fan, Z.; Wu, Z.; Dai, X.; Huang, S.; Chen, J. Target-oriented opinion words extraction with target-fused neural sequence labeling. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; Long and Short Papers; Volume 1, pp. 2509–2518.

31.	Pontiki, M.; Galanis, D.; Papageorgiou, H.; Androutsopoulos, I.; Manhar, S.; Al-Smadi, M.; Al-Ayyoub, M.; Zhao, Y.; Qin, B.; De Clercq, O.; et al. Orphée de Clercq Semeval-2016 task 5: Aspect based sentiment analysis. In Proceedings of the International Workshop on Semantic Evaluation, San Diego, CA, USA, 16–17 June 2016; pp. 19–30.

32.	Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.

33.	Zhao, H.; Huang, L.; Zhang, R.; Lu, Q.; Xue, H. 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, Online, 5–10 July 2020; pp. 3239–3248.

34.	Zhang, W.; Deng, Y.; Li, X.; Yuan, Y.; Bing, L.; Lam, W. Aspect Sentiment Quad Prediction as Paraphrase Generation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, Online, 7–11 November 2021; pp. 9209–9219.

35.	Wang, W.; Pan, S.J.; Dahlmeier, D.; Xiao, X. Coupled multi-layer attentions for co-extraction of aspect and opinion terms. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA,4–9 February 2017; Volume 31.

36.	Dai, H.; Song, Y. Neural Aspect and Opinion Term Extraction with Mined Rules as Weak Supervision. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5268–5277.

37.	Li, X.; Bing, L.; Li, P.; Lam, W. A unified model for opinion target extraction and target sentiment prediction. In Proceedings of the AAAI conference on artificial intelligence, Honolulu , HI, USA, 27 January–1 February 2019; Volume 33, pp. 6714–6721.

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

39.	Bekoulis, G.; Deleu, J.; Demeester, T.; Develder, C. Joint entity recognition and relation extraction as a multi-head selection problem. *Expert Syst. Appl.* **2018**, *114*, 34–45. [CrossRef]