

Received 29 August 2022, accepted 18 September 2022, date of publication 26 September 2022, date of current version 4 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3210180

RESEARCH ARTICLE

A Multi-Task Dual-Encoder Framework for Aspect Sentiment Triplet Extraction

HAI HUAN¹, (Member, IEEE), ZICHEN HE¹, YAQIN XIE², AND ZELIN GUO²

¹School of Artificial Intelligence (School of Future Technology), Nanjing University of Information Science and Technology, Nanjing 210044, China

²School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

Corresponding author: Hai Huan (haihuan@nuist.edu.cn)

This work was supported by the National Natural Science Foundation of China (NSFC) under Grant 62001238.

ABSTRACT Aspect Sentiment Triplet Extraction (ASTE) is a complex and important task in aspect-based sentiment analysis task, which aims to extract aspect-sentiment-opinion triplets from review sentences, to acquire comprehensive information for sentiment analysis. Most of the existing methods use pipeline approaches or end-to-end sequence tagging approaches to solve the ASTE task. However, the pipeline approaches suffer from error accumulation in practical applications. The existing sequence tagging approaches ignore the feature information of the three elements themselves, and cannot model and infer the three elements effectively by placing each word in the same position as importance. Based on this, a multi-task dual-encoder framework is proposed. First, a dual-encoder is constructed to encode and fuse sentence information and semantic information, respectively. Then, the signs and constraints implied between word pairs are used to complete multi-task inference and triplet decoding. Meanwhile, two grid tagging methods and their corresponding inference strategies are designed for the multi-task. The auxiliary task is used as a regularization of the main task, which improves the correct inference ability of the inference strategy for the main task and the robustness of the framework. Extensive testing on two benchmark datasets shows that the proposed framework is simple and effective, and significantly outperforms the existing methods.

INDEX TERMS Aspect based sentiment analysis, aspect sentiment triplet extraction, natural language processing.

I. INTRODUCTION

With the rapid development of information technology, more people tend to post reviews on the InterNet, and these reviews are often used as important reference information for people to make decisions. Therefore, the automatic analysis of opinions and sentiments expressed in review sentences has become an important research direction in the field of Natural Language Processing (NLP). According to the analysis granularity, text sentiment analyses can be divided into document-level sentiment analysis, sentence-level sentiment analysis, and aspect-based sentiment analysis. Among them, Aspect-Based Sentiment Analysis (ABSA) [1], [2], [3] is the most complex, because it needs to understand the semantic information at the word and sentence levels, which requires not only the analysis of explicit language expression structure

but also an understanding of implicit semantic expression. The ABSA task consists of three basic subtasks: Aspect Term Extraction (ATE) [4], [5], [6], Opinion Term Extraction (OTE), [7], [8], and Aspect Sentiment Classification (ASC) [9], [10], [11], [12], [13], [14], [15], [16]. Based on these three subtasks, new subtasks are formed by combining them in pairs. ATE and OTE are combined to form Aspect-Opinion Pair Extraction (AOPE) [17], [18]. ATE and ASC are integrated to form Aspect-Polarity Co-Extraction (APCE) [19], [20]. Recently, attempts to combine the three basic subtasks into a new subtask have been made, leading to the Aspect Sentiment Triplet Extraction (ASTE) task [21], [22], [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34]. This task requires extracting aspect-sentiment-opinion triplets from a given review sentence. As shown in Figure 1, the review sentence contains two triplets, (*cake with truffles*, *try*, *positive*) and (*cake with truffles*, *incredible*, *positive*), where “*cake with truffles*” is the aspect term of

The associate editor coordinating the review of this manuscript and approving it for publication was Guangjie Han¹.

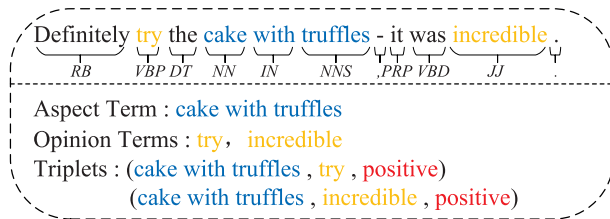


FIGURE 1. An example of the ASTE task. The spans shown in blue are aspect terms. Spans shown in orange are opinion terms, and sentiment polarity is shown in red.

the review sentence, “try” and “incredible” are opinion terms, that is, the emotional expression of the aspect term, “positive” means that the sentiment polarity of the aspect term is positive. Through such extraction, the various emotional factors contained in the review sentence can be clearly known, making it more suitable for application in practical scenarios.

The ASTE task can be roughly divided into two technical routes. Route one adopts two-stage approaches [21], [30]. [21] extracted the potential aspect terms, sentiment polarity, and opinion terms in the first stage; in the second stage, a Multilayer Perceptron (MLP) was used to pair each aspect term with an opinion term to form a complete triplet. However, this method is computationally inefficient, does not consider the relationships and interactions between individual word pairs, and requires training of three separate frameworks, making it vulnerable to error propagation. Route two employs end-to-end sequence tagging approaches [22], [23], [24], [26], [31], [32], [33], [34]. Previous research [23] proposed a Grid Tagging Scheme (GTS) that can tag the relationship between all word pairs in a grid. However, this approach places each word of a term in an equally important position, which cannot be effectively modeled between aspect terms and opinion terms and is not conducive to extracting aspect terms and opinion terms composed of multiple words. This approach also ignores the global semantics [19] and the part-of-speech features of words that constitute aspect terms and opinion terms. In practice, when extracting triplets, there are one-to-many and many-to-one situations between aspect terms and opinion terms, which further increases the difficulty of the ASTE task.

Based on the above analysis, we propose a multi-task dual-encoder framework to extract triplets in one go in an end-to-end fashion. Our framework uses both the token BERT encoder [35] and the Label Name BERT encoder. The former encodes the token of the review sentence, and the latter is inspired by previous work [36], which encodes the part-of-speech feature of the words in the sentence. Then, two grid tagging tasks are designed, one of which is used as a main task (called the word pair relationship tagging main task) and is used to tag all word pair relations, including aspect terms, opinion terms, and sentiment polarity; the other is used as an auxiliary task (called the boundary prediction auxiliary task) and is used to predict the boundaries of each aspect and

opinion terms (i.e., the start and end positions of aspect terms and opinion terms). At the same time, we define the extraction center (i.e., the beginning words of aspect terms and opinion terms) and the sentiment center (i.e., the end words of aspect terms) according to the position of different words in the terms, and design effective tagging methods and inference strategies for the above two tasks. Since these two tasks are jointly addressed in training, the auxiliary task can be regarded as a regularization of the main task to strengthen the extraction of aspect terms and opinion terms. The framework integrates the part-of-speech features and contextual features of the review sentence, and after the joint training of multi-task, only one inference on the datasets is needed to extract the triplet from the final prediction.

Extensive experiments are conducted on two benchmark datasets and compared with existing approaches. Experimental results show that the framework proposed in this paper is significantly better than existing approaches, and further research shows that each component we proposed is both simple and effective. Specifically, the main contributions of this paper are drawn as follows:

- A novel encoder is proposed to obtain the semantic features of the words themselves, providing additional information for the representation of each word, further enhancing the contextual information of the current sentence, and sufficiently utilizing the token-level and sentence-level semantic information.
- A multi-task grid tagging framework is proposed, which considers the relative positions of words in terms and assigns different importance to them. Based on this, two grid tagging methods and inference strategies are proposed to further enhance the ability of the framework to extract triplets.
- Extensive experimental results show that the proposed simple framework achieves significantly better extraction results than existing complex approaches on two benchmark datasets.

II. RELATED WORK

A. SINGLE EXTRACTION IN ASPECT-BASED SENTIMENT ANALYSIS

The ABSA task consists of three basic single extraction subtasks: ATE (Aspect Term Extraction), OTE (Opinion Term Extraction), and ASC (Aspect Sentiment Classification). Among them, the purpose of the ATE task is to extract aspect terms from sentences, which is usually regarded as a sequence labelling task. [4] employed two types of pre-trained embeddings (general-purpose embeddings and domain-specific embeddings) to represent sentences and then used a simple convolutional neural network to achieve good results. [5] formalized the ATE task as a sequence-to-sequence (Seq2Seq) learning task and introduced gated unit networks and position-aware attention to improve the model’s ability to extract the aspect terms. The OTE task is generally regarded as an auxiliary task of the ATE task,

and its purpose is to extract opinion terms of a given aspect term. To this end, [7] proposed to incorporate the syntactic structure of sentences and the syntax-based opinion possibility score into the OTE task. The ASC task aims to classify the sentiment polarity of the aspect terms. In [9], a long short-term memory neural network was proposed based on an attention mechanism. It can pay attention to the connection between aspect terms and the whole sentence and focus on the important parts. Elsewhere, a syntactic dependency relational graph attention network [16] is used to exploit syntactic dependency tree information. However, in these studies, none of the extracted aspect terms and opinion terms were paired.

B. JOINT EXTRACTION IN ASPECT-BASED SENTIMENT ANALYSIS

Recently, some researchers have focused on more complex subtasks, they combine the basic subtasks in ABSA to form new subtasks. These subtasks mainly include AOPE (Aspect-Opinion Pair Extraction) and APCE (Aspect-Polarity Co-Extraction). The AOPE task aims to extract both aspect terms and opinion terms from the review sentences. Unlike the OTE task, the AOPE task does not give aspect terms and requires a joint extraction from sentences. To this end, [18] proposed a synchronous double-channel recurrent network, which took the opinion term extraction unit and relation detection unit as two channels to extract opinion terms and relations simultaneously. The APCE task requires the extraction of both aspect terms and sentiment polarity from sentences. To this end, [19] proposed an interactive multi-task learning network that introduces a message passing architecture where information is iteratively passed to different tasks through a shared set of latent variables.

None of the above tasks can be used to extract the full sentiment expression of a sentence. To address this problem, the ASTE task was first proposed in [21] to extract aspect-sentiment-opinion triplets from review sentences; they adopted a pipeline approach to solving the problem. In the first stage, the potential aspect terms and their sentiment polarity and opinion terms are extracted; in the second stage, the aspect terms and opinion terms are matched to form a complete triplet. Reference [24] designed a new position-aware tagging scheme, which can jointly extract triplets in an end-to-end fashion. Thereafter, [32] utilized a graph neural network based on another work [23] to encode sentences, and then used the grid tagging scheme [23] to extract the triplets. References [27] and [29] described the ASTE task as a Machine Reading Comprehension (MRC) problem. The former extracted triplets by joint training two BERT-MRC frameworks with shared parameters, and the latter adopted a Bidirectional-MRC (BMRC) structure for multi-turn machine reading. Recently, [33] proposed a hierarchical reinforcement learning framework that treats aspect terms and opinion terms as parameters of sentiment polarity, considering the interactions among triplets while improving efficiency. In another work, a span-based approach [25],

which considers the interactions between the aspect term and opinion term spans, was adopted; this approach uses the semantics of the entire span to predict sentiment polarity. Although these tasks have achieved good performance on the ASTE task, they do not incorporate the feature information of sentences themselves, and it is difficult to extract aspect terms and opinion terms composed of multiple words. By contrast, the framework we proposed adopts two improved grid tagging schemes, which are more sensitive to aspect terms and opinion terms composed of multiple words, it can also extract complete triplets at one time, and is unaffected by the position between aspect terms and opinion terms.

III. APPROACH

In this section, the ASTE task is first defined. Next, our framework consisting of a dual-encoder and a multi-task framework is described. Subsequently, the inference strategies and triplet decoding algorithm are introduced. Finally, how to train the framework is demonstrated.

A. TASK DEFINITION

Given a review sentence $X = \{w_1, w_2, \dots, w_n\}$, it contains n words, the goal of the ASTE task is to extract all triplets $T = \{(a, o, s)_m\}_{m=1}^{|P|}$ from the sentence X , where the notations a and o respectively denote an aspect term and an opinion term, and s is the sentiment polarity corresponding to the aspect term, $s \in \{POS, NEU, NEG\}$. $|P|$ represents the total number of triplets in the sentence.

B. BERT

Inspired by the successful practice on many NLP tasks, we use BERT as our backbone model to encode the context information. In order to better understand the structure used in subsequent articles, we briefly introduce BERT before introducing the framework.

Recall that BERT is a language model based on a multi-layer bidirectional Transformer [37] proposed by Google. The pre-trained BERT can generate word vectors of the sequence, which can be used as a high-quality input for downstream tasks. Specifically, we transform the sentence into "[CLS]+sentence+[SEP]" to represent the entire input. The special tokens [CLS] and [SEP] represent the beginning token and segment token, respectively. The processed sequence is then fed into BERT for context encoding. First, each token is converted into a vector by summing its TOKEN EMBEDDING, SEGMENT EMBEDDING, and POSITION EMBEDDING. After that, the vector sequence is fed into a stack of Transformer layers to obtain the encoded contextual information. We use the hidden layer output of the last Transformer block as the context representation. For ease of understanding, the hidden representation in the subsequent article refers to the representation of the word in each sentence and does not contain special tokens. The details of the BERT architecture used in this study are explained in Section IV-B.

C. FRAMEWORK

The overall architecture of the proposed multi-task dual-encoder framework is shown in Figure 2. In the encoding stage, a double BERT encoder is adopted to encode the token feature of the sentence and the part-of-speech feature of the word itself, and the two feature representations are then added for feature fusion to obtain the representation of the review sentence. Multi-task joint training is then applied to the representation to predict word pair relations and the boundaries of aspect terms and opinion terms, respectively. Meanwhile, two efficient inference strategies are proposed to exploit the latent signs and constraints between word pairs. In the decoding stage, the word pair relations tagging main task grid, which can extract complete triplets from the input sentence in one go in an end-to-end fashion, can be decoded.

1) CONTEXT ENCODING WITH DUAL-ENCODER

In this section, we introduce the dual-encoder structure, which sufficiently utilizes token-level and sentence-level semantic information.

a: TOKEN BERT ENCODER

The Token BERT encoder generates a contextual representation $\hat{\mathbf{H}} = \{\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \dots, \hat{\mathbf{h}}_n\}$ for a given sentence X . Since BERT mainly uses POSITION EMBEDDING to simply encode the position of words, it cannot guarantee that continuous sequence information is captured every time. Therefore, to capture continuous sequence information and global semantic information in the sentence, and enhance the context-awareness of the model for the current review sentence, our framework also adds a layer of Bidirectional Long Short-Term Memory (BiLSTM) [38] after using the Token BERT encoder to obtain the representation $\tilde{\mathbf{H}} = \{\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \dots, \tilde{\mathbf{h}}_n\}$.

$$\tilde{\mathbf{H}} = [\overrightarrow{LSTM}(\hat{\mathbf{H}}); \overleftarrow{LSTM}(\hat{\mathbf{H}})], \quad (1)$$

where $\overrightarrow{LSTM}(\hat{\mathbf{H}})$ and $\overleftarrow{LSTM}(\hat{\mathbf{H}})$ represent the representation of BiLSTM forward propagation and backpropagation respectively, and $[\cdot; \cdot]$ denotes the vector concatenation operation.

b: LABEL NAME BERT ENCODER

The name of the label carries the meaning information of the label, and this information can also be summarized by the model from the data. Therefore, we use Stanza¹ [39] to parse the abbreviation of the part-of-speech label of each word. Then the abbreviation of the part-of-speech label is assigned to a specific meaning and converted into the corresponding label name. For example, the part-of-speech abbreviations of each word in the sentence in Figure 2 are RB, VBP, DT, NN, IN, and NNS. According to the conversion table, these abbreviations are converted into specific words and then form a new sentence “other verb qualifier noun other noun”. Next, the new sentence is sent to the Label Name BERT encoder to

TABLE 1. The label name conversion table.

Label	Converted label name
CC	conjunction
CD	number
DT	determiner (We use “qualifier” instead.)
PRP	personal
JJ,JJS	adjective
NN,NNS,NNP,NNPS	noun
VB,VBZ,VBD,VBP, VBN,VBG	verb
OTHER LABEL	All labels except the preceding ones are replaced with “other”.

get the encoding representation $\bar{\mathbf{H}} = \{\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2, \dots, \bar{\mathbf{h}}_n\}$. The label name conversion table is shown in Table 1.

Finally, by adding the two feature representations for feature fusion, the representation \mathbf{h}_i of each word, the representation \mathbf{H} of the whole review sentence, and the representation \mathbf{r}_{ij} of each word pair (w_i, w_j) are obtained as follows:

$$\mathbf{h}_i = \tilde{\mathbf{h}}_i + \bar{\mathbf{h}}_i, \quad (2)$$

$$\mathbf{H} = \tilde{\mathbf{H}} + \bar{\mathbf{H}}, \quad (3)$$

$$\mathbf{r}_{ij} = [\mathbf{h}_i; \mathbf{h}_j], \quad (4)$$

where $[\cdot; \cdot]$ represents the vector concatenation operation.

2) MULTI-TASK JOINT TRAINING FRAMEWORK

To focus the framework’s attention on those influential feature representations, we propose word pair relationship tagging as the main task and boundary prediction as an auxiliary task. The schematic diagram of multi-task training is shown in Figure 3. The two tasks jointly train the underlying parameters (if the inference stage is included, the underlying parameters are still shared by the two tasks during inference).

a: THE MAIN TASK FRAMEWORK

The purpose of the main task is to tag aspect terms, opinion terms, and sentiment polarity corresponding to the aspect terms. Reference [23] used six tags $G = \{N, A, O, NEG, NEU, POS\}$ to represent the relations of each word pair (w_i, w_j) in the sentence. In contrast, a ten-tag tagging method, which uses a unified tag with position information $Z = \{N1, A-B, A-I, A-A, O-B, O-I, O-O, NEG, NEU, POS\}$ to represent the relationships of each word pair (w_i, w_j) in the sentence, is proposed in this study, where $-B$ and $-I$ represent the relative position relationship of the word pair, indicating the beginning or the inner part of a term. $-A$ and $-O$ are used to detect whether a word pair formed by two different words belongs to the same aspect or opinion term, respectively. The ten-tag tagging method will help our model to infer more accurately and extract the final triplets (the inference strategies are described in Section III-C3). The specific meanings of these tags are listed in Table 2. Figure 3 (left) shows an example of the main task tags for the sentence in Figure 1. The main task uses an upper triangular grid to accelerate the tagging of word pair relationships.

¹<https://stanfordnlp.github.io/stanza/>

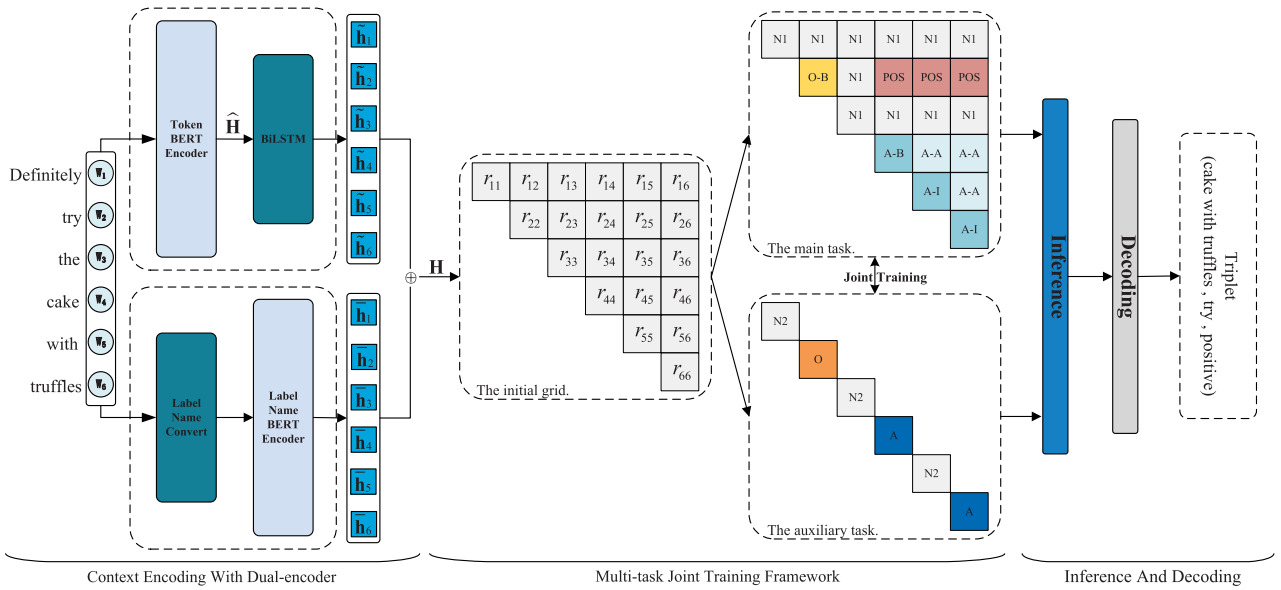


FIGURE 2. The architecture of our proposed multi-task dual-encoder framework.

TABLE 2. The meanings of tags for the main task.

Tags	The meanings of Tags
N1	no relations for word pair (w_i, w_j) .
A-B	the word pair (w_i, w_j) belongs to the beginning of an aspect term.
A-I	the word pair (w_i, w_j) belongs to the inside of an aspect term.
A-A	the word pair (w_i, w_j) belongs to the same aspect term.
O-B	the word pair (w_i, w_j) belongs to the beginning of an opinion term.
O-I	the word pair (w_i, w_j) belongs to the inside of an opinion term.
O-O	The word pair (w_i, w_j) belongs to the same opinion term.
NEG	the sentiment polarity of the word pair (w_i, w_j) is negative.
NEU	the sentiment polarity of the word pair (w_i, w_j) is neutral.
POS	the sentiment polarity of the word pair (w_i, w_j) is positive.

TABLE 3. The meanings of tags for the auxiliary task.

Tags	The meanings of Tags
N2	the word pair (w_i, w_j) does not belongs to an aspect term or an opinion term.
A	the word pair (w_i, w_j) belongs to the start or end of an aspect term.
O	the word pair (w_i, w_j) belongs to the start or end of an opinion term.

b: THE AUXILIARY TASK FRAMEWORK

The purpose of the auxiliary task is to tag the boundaries of the aspect terms and opinion terms in each sentence. Previous research [34] has proved that the target boundary tags are beneficial to the prediction of unified tags. In this study, designing an auxiliary task as a regularization term can improve the sensitivity and discriminative ability of our framework to aspect terms and opinion terms composed of multiple words. It can also give additional penalties for mislabeled aspect

terms and opinion terms, thus prompting the main task to pay more attention to mislabeled word pairs. The auxiliary task also adopts a grid tagging scheme and uses unified tags $Q = \{N2, A, O\}$ to ascertain whether the word pair (w_i, w_j) belongs to the boundary of an aspect term or an opinion term. In this task, the spans of aspect terms and opinion terms do not overlap, so we only use a single A or O to mark the boundaries of aspect terms or opinion terms. The specific meanings of these tags are listed in Table 3. Figure 3 (right) shows an example of auxiliary task tags for the sentences in Figure 1. The boundary prediction auxiliary task only fills the diagonal grid.

3) INFERENCE AND DECODING

a: INFERENCE STRATEGIES

Herein the heuristic inference strategies adopted by the framework are introduced. As mentioned in Section III-C2, a ten-tag tagging method is introduced, which helps the framework to infer. For an aspect term, no matter how many words it consists of, the first word pair (w_i, w_i) must be predicted as $A-B$ to be correct. Therefore, we think that the beginning of the terms is the extraction center and has an important position. In addition, if the predicted tag of the word pair (w_{i+1}, w_{i+1}) is not $A-I$, which means that the aspect term is composed of only one word. If the predicted tag of the word pair (w_{i+1}, w_{i+1}) is $A-I$, and only the tag of the cross grid word pair (w_i, w_{i+1}) is predicted as $A-A$, the continuous span from word pair (w_i, w_i) to word pair (w_{i+1}, w_{i+1}) can be determined as an aspect term.

As shown in Figure 3 (left), “cake with truffles” is an aspect term consisting of three words, and the corresponding terms (cake, cake), (with, with), and (truffles, truffles) on the main diagonal are marked as $A-B$ or $A-I$. According to

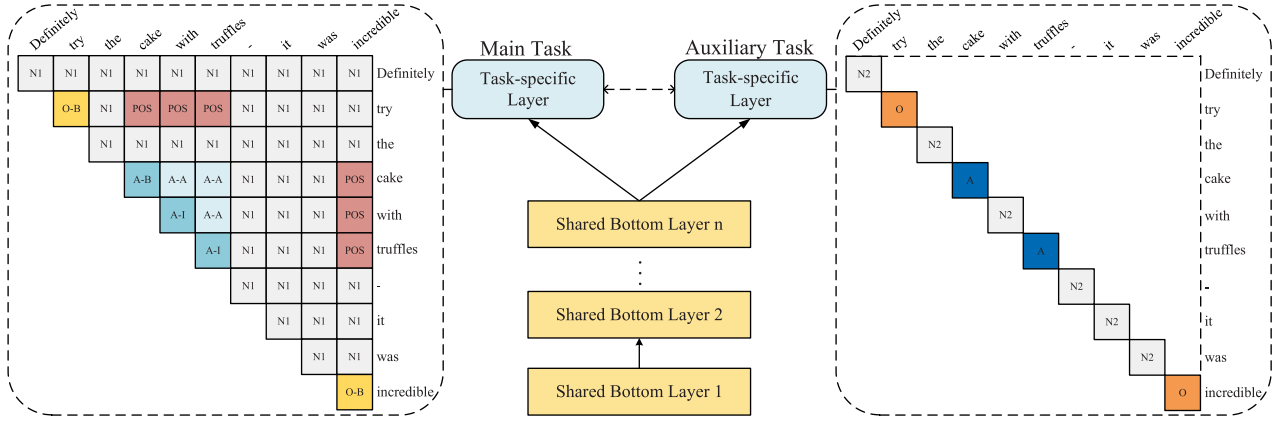


FIGURE 3. An example of the multi-task joint training. The left is the word pair relationship tagging main task and the right is the boundary prediction auxiliary task.

the inference strategy, if it is only based on the existence of continuous spans, the continuous span from “cake” to “truffles” cannot be determined as an aspect term. For the grid where word pairs intersect, such as (cake, with), it must also be marked as A-A to satisfy the judgment conditions we proposed. The inference strategy simplifies the matching constraints on the complete aspect terms composed of a single word pair and strengthens the matching constraints and judgment conditions for the complete aspect terms composed of multiple word pairs. (The same is true for predictions of opinion terms.)

Furthermore, if one word pair containing w_i is predicted as A-B or A-I, other word pairs containing w_i are less likely to be predicted as O-B or O-I, and vice versa. In other words, the relationship between w_i and other words help the main task to infer the tag of the word pair (w_i, w_j). Based on the above, it can be known that the ten-tag tagging method further constrains the possible range of the predicted tag in the next round. To take advantage of these latent signs and constraints, the following inference strategy is used to obtain the new feature representation \mathbf{z}_{ij}^t and predicted probability distribution \mathbf{p}_{ij}^t of the word pair (w_i, w_j):

$$\mathbf{p}_i^{t-1} = \text{maxpooling}(\mathbf{p}_{i,:}^{t-1}), \quad (5)$$

$$\mathbf{p}_j^{t-1} = \text{maxpooling}(\mathbf{p}_{:,j}^{t-1}), \quad (6)$$

$$\mathbf{q}_{ij}^{t-1} = [\mathbf{z}_{ij}^{t-1}; \mathbf{p}_i^{t-1}; \mathbf{p}_j^{t-1}; \mathbf{p}_{ij}^{t-1}], \quad (7)$$

$$\mathbf{z}_{ij}^t = \mathbf{W}_q \mathbf{q}_{ij}^{t-1} + \mathbf{b}_q, \quad (8)$$

$$\mathbf{p}_{ij}^t = \text{softmax}(\mathbf{W}_s \mathbf{z}_{ij}^t + \mathbf{b}_s), \quad (9)$$

In the above process, $\mathbf{p}_{i,:}^{t-1}$ denotes all predicted probabilities between words w_i and other words. Equations (5), (6), and (7) are used to observe the characteristics of the probability distribution of each word pair itself and between word pairs. t refers to the t -th turn. \mathbf{W}_q , \mathbf{b}_q , \mathbf{W}_s and \mathbf{b}_s are trainable parameters. Because an upper triangular grid is used,

so $\mathbf{p}_{i,:}^{t-1} = (\mathbf{p}_{1:i,i}^{t-1}, \mathbf{p}_{i,i:n}^{t-1})$. The initially predicted probability \mathbf{p}_{ij}^0 and feature representation \mathbf{z}_{ij}^0 of the word pair (w_i, w_j) are set as follows:

$$\mathbf{p}_{ij}^0 = \text{softmax}(\mathbf{W}_s \mathbf{r}_{ij} + \mathbf{b}_s), \quad (10)$$

$$\mathbf{z}_{ij}^0 = \mathbf{r}_{ij}, \quad (11)$$

Finally, the last round of predictions \mathbf{p}_{ij}^L can be used to extract triplets according to Algorithm 1.

For the auxiliary task, there are also some potential signs and constraints to help the framework make inferences. When the first A tag is found from top to bottom along the main diagonal, we continue to search backward. If the A tag is encountered again, it is considered that the first A tag and the second A tag are the start and end positions of this aspect term, respectively. If the O tag is encountered, the first A tag is considered to be the boundary of this aspect term alone, that is, this aspect term consists of a word. Similarly, when the first one found is an O tag, the inference can be as outlined above. In addition, for aspect terms and opinion terms composed of multiple words, the A tag or O tag single encountered is always the start of the aspect or opinion term, and the A tag or O tag double encountered is always the end of the aspect term or opinion term, and it is always paired with the nearest preceding A or O tag to form an aspect term span or an opinion term span.

As shown in Figure 3 (right), when inferring from top to bottom, the O tag of “try” is encountered first, and then the A tag of “cake” is encountered, which indicates that “try” is an opinion term composed of a single word. And then, when the A tag of “truffles” is encountered, because the previous “cake” is also an A tag, “cake” and “truffles” are the boundaries of this aspect term. The prediction formula calculation of the auxiliary task is consistent with the main task. Because the auxiliary task only uses the main diagonal grid, in the auxiliary task, $\mathbf{p}_{i,:}^{t-1} = (\mathbf{p}_{i,i}^{t-1}, \mathbf{p}_{i,i}^{t-1})$.

Compared with other complex frameworks, our framework requires only one inference to achieve superior performance.

b: DECODING ALGORITHM

Since our purpose is to extract aspect sentiment triplets, only the grid of main task needs to be decoded. The decoding algorithm is shown in Algorithm 1. First, we use the predicted tags of all word pairs (w_i, w_j) on the main diagonal to decode the possible aspect terms and opinion terms, as shown in lines 2 to 13 in Algorithm 1, consisting of a single $A-B$ tag or from $A-B$ tag to $A-I$ tag continuous span (and its intersection is marked as $A-A$) will be identified as an aspect term. The same is true for the identification of opinion terms. Second, it is judged whether the extracted aspect terms and opinion terms match the pair. For aspect term a and opinion term o , the prediction relationship s of the word pair (w_i, w_j) is calculated, where $w_i \in a, w_j \in o$ or $w_i \in o, w_j \in a$. If $s \in \{POS, NEU, NEG\}$, the aspect term and the opinion term are deemed to be paired. Otherwise, the two terms are not paired. Finally, as in lines 16 to 17 in Algorithm 1, the most predicted sentiment tag s is regarded as the sentiment polarity of the aspect term. It should be noted that if the same number of different polarities appear, the predicted tag of the last word pair of the aspect term is chosen as the sentiment polarity of the aspect term, because usually the last word in the aspect term is the sentiment center of the whole term.

D. MODEL TRAINING

For the main task, the loss function is defined as the cross-entropy loss function between the ground-truth distribution and the predicted tagging distribution \mathbf{p}_{ij}^L of all word pairs:

$$L_t = - \sum_{i=1}^n \sum_{j=i}^n \sum_{k \in Z} \prod (y_{ij} = k) \log(\mathbf{p}_{i,j|k}^L), \quad (12)$$

where y_{ij} represents the ground truth tag of the word pair (w_i, w_j) , $\prod(\cdot)$ is the indicator function, and Z denotes the tag set, in the main task, $Z = \{N1, A-B, A-I, A-A, O-B, O-I, O-O, NEG, NEU, POS\}$. Herein, when $y_{ij} \in \{A-B, O-B\}$, we increase the loss weight because the beginning of the terms is defined as the extraction center by us.

Similarly, for the auxiliary task, the loss function is defined as:

$$L_l = - \sum_{i=1}^n \sum_{j=i}^i \sum_{k' \in Q} \prod (y_{ij} = k') \log(\mathbf{p}_{i,j|k'}^L), \quad (13)$$

where y_{ij} represents the ground truth tag of the word pair (w_i, w_j) , $\prod(\cdot)$ is the indicator function. Q denotes the tag set, in the auxiliary task, $Q = \{N2, A, O\}$.

Finally, multi-task joint training is performed, and the overall loss function is as follows:

$$L = L_t + \alpha L_l. \quad (14)$$

Among them, the hyperparameter α can be used to adjust the influence of the auxiliary task on the loss of the main task. The selection of the hyperparameter α is discussed in Section V-D.

Algorithm 1 Decoding Algorithm for the ASTE Main Task

Input: The predicted results P of a sentence X . $P(w_i, w_j)$ denotes the predicted tag of the word pair (w_i, w_j) .

Output: An aspect sentiment triplet set T_{set} of the given sentence.

```

1: Initialize the aspect term set  $A_{set}$ , opinion term set  $O_{set}$ ,
   and aspect sentiment triplet set  $T_{set}$  with  $\emptyset$ .
2: while a span left index  $l \leq n$  and right index  $r \leq n$  do
3:   if  $P(w_l, w_l) = A-B$  when  $l \leq i \leq r$ , meanwhile
      $P(w_{i+1}, w_{i+1}) \neq A-I$  then
4:     Regard the word  $w_i$  as an aspect term  $a$ ,  $A_{set} \leftarrow$ 
        $A_{set} \cup \{a\}$ 
5:   else if  $P(w_l, w_l) = A-B$  when  $l \leq i \leq r$ , mean-
     while  $P(w_r, w_r) = A-I$  and  $P(w_{r+1}, w_{r+1}) \neq A-I$  and
      $P(w_{i:r,i+1}, w_{i+1,i+1:r}) = A-A$  then
6:     Regard the words  $\{w_i, \dots, w_r\}$  as an aspect term
        $a$ ,  $A_{set} \leftarrow A_{set} \cup \{a\}$ 
7:   end if
8:   if  $P(w_l, w_l) = O-B$  when  $l \leq i \leq r$ , meanwhile
      $P(w_{i+1}, w_{i+1}) \neq O-I$  then
9:     Regard the word  $w_i$  as an opinion term  $o$ ,  $O_{set} \leftarrow$ 
        $O_{set} \cup \{o\}$ 
10:  else if  $P(w_l, w_l) = O-B$  when  $l \leq i \leq r$ , mean-
    while  $P(w_r, w_r) = O-I$  and  $P(w_{r+1}, w_{r+1}) \neq O-I$  and
     $P(w_{i:r,i+1}, w_{i+1,i+1:r}) = O-O$  then
11:    Regard the words  $\{w_i, \dots, w_r\}$  as an opinion term
       $o$ ,  $O_{set} \leftarrow O_{set} \cup \{o\}$ 
12:  end if
13: end while
14: while  $a \in A_{set}$  and  $o \in O_{set}$  do
15:   while  $w_i \in a$  and  $w_j \in o$  or  $w_i \in o$  and  $w_j \in a$  do
16:    if the most predicted sentiment tag  $s \in$ 
       $\{POS, NEU, NEG\}$  then
17:       $T_{set} \leftarrow T_{set} \cup \{(a, o, s)\}$ 
18:    end if
19:  end while
20: end while
21: return the set  $T_{set}$ 

```

IV. EXPERIMENTS

In this section, the details of the experiments are introduced first. Including the datasets, experimental settings, and evaluation method. It then briefly introduces the compared methods, and finally presents the experimental results and discusses them.

A. DATASETS

We evaluate our framework on two benchmark datasets: ASTE-Data-V1 [21] and ASTE-Data-V2 [24]. All review sentences are derived from SemEval Challenges [40], [41], [42]. The ASTE-Data-V2 datasets are refined from the previous ASTE-Data-V1 datasets.² The statistics of these two

²<https://github.com/xuuluuuu/SemEval-Triplet-data>

TABLE 4. Statistics of the ASTE-Data-V1 datasets.

Dataset	14res		14lap		15res		16res	
	#S	#T	#S	#T	#S	#T	#S	#T
train	1300	2145	920	1265	593	923	842	1289
dev	323	524	228	337	148	238	210	346
test	496	862	339	490	318	455	320	465

TABLE 5. Statistics of the ASTE-Data-V2 datasets.

Dataset	14res		14lap		15res		16res	
	#S	#T	#S	#T	#S	#T	#S	#T
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

datasets can be seen in Tables 4 and 5, where #S represents the total number of review sentences in the datasets, and #T denotes how many triplets exist in the datasets, “res” and “lap” indicate that the datasets are from the restaurant domain and the laptop domain.

B. EXPERIMENTAL SETTINGS

In the present work, we adopt the uncased base version of BERT [35], which contains 110M parameters. During training, we use AdamW [43] to optimize parameters. The fine-tuning rate and learning rate of BERT are set to $2e-5$ and $1e-3$, respectively, and the hidden vector dimension is 768. The learning rate of other trainable parameters is $1e-3$, and the dropout rate [44] is set to 0.5. The hyperparameter α of the auxiliary task is set to 0.01. The best model parameters are selected based on the best F1-score on the validation sets to evaluate the testsets and the average results of five times with different random seeds are reported.

C. EVALUATION

To assess the comprehensive performance of our framework, in addition to the ATSE task, three other subtasks (including ATE, OTE, and AOPE) are also evaluated and the related results are discussed in section V-B. Our experiments use precision (P), recall (R), and F1-score as evaluation metrics, where the F1-score is used to comprehensively assess the performance of the framework. The correct triplets are considered to have been extracted only when the predicted aspect term span, opinion term span, and sentiment polarity conform to the ground truth.

D. COMPARED METHODS

Our framework is compared with the following approaches for the ASTE task:

1) Pipeline methods:

- Peng-two-stage: In the first stage, [21] extracted the potential aspect terms and their sentiment polarity and opinion terms. In the second stage, the aspect terms and opinion terms were paired to output the complete triplets.

- Li-unified-R+: [21] first used the Li-unified opinion enhancement module proposed in [34] to determine the aspect terms, sentiment polarity, and opinion terms. Then, the relationship classifier proposed in [21] was then adopted to conduct relationship matching.

2) MRC methods:

- BMRC: [29] transformed the ASTE task into a multi-turn machine reading comprehension task. In this task, three types of queries and a two-way MRC structure were designed. One direction identifies aspect terms, opinion terms, and sentiment polarity in turn, and the other direction first identifies opinion terms and then aspect terms.
- Dual-MRC: Dual-MRC was proposed in [27] to solve the ASTE task by jointly training two BERT-MRC frameworks with shared parameters.

3) End-to-end methods:

- JET-T: [24] proposed the first end-to-end framework and a new position-aware tagging method. JET-T incorporates the features of opinion terms and sentiment polarity into aspect terms, and uses the LSTM layer and Conditional Random Field (CRF) layer to predict the label of each word.
- JET-O: A variant of JET-T proposed by [24] that incorporates features of aspect terms and sentiment polarity into opinion terms, and then uses the same way as JET-T to predict labels and capture elements in triplets.
- GTS-BERT: The Grid Tagging Scheme (GTS) is a tagging method proposed in [23], which only needs a unified grid to handle the ASTE task in an end-to-end fashion.
- S³E²: [32] designed a graph sequence dual representation and modelling paradigm for the ASTE task, which uses graphs to learn and represent the semantic and syntactic relationships between word pairs in sentences, and uses a graph neural network to encode them to extract triplets.
- EIN: [26] adopted two encoders to display interaction bi-directionally, and used a multi-layer sequence encoder for target-opinion detection and sentiment polarity classification simultaneously.
- ASTE-RL: In the framework of hierarchical reinforcement learning, [33] considered aspect terms and opinion terms as sentiment expression parameters, and considered the interaction between triplets, which improved the efficiency and enabled the model to deal with multiple triplets.
- UniASTE: [31] first used sequence tagging to predict the boundaries information of opinion targets and opinion expressions, and then introduced a target-aware tagging scheme, taking each word in the sentence as a potential opinion target in turn.

TABLE 6. Experimental results (%) of the ASTE task on the ASTE-Data-V1 datasets. The baseline results with † are from [24]. Other baseline results are retrieved from the original papers. If the baseline to be compared has a Bert-based version, the BERT-based version is selected for the results. The best scores are marked in bold.

Models	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Li-unified-R+†	41.44	68.79	51.68	42.25	42.78	42.47	43.34	50.73	46.69	38.19	53.47	44.51
Peng-two-stage†	44.18	62.99	51.89	40.40	47.24	43.50	40.97	54.68	46.79	46.76	62.97	53.62
JET-T	70.20	53.02	60.41	51.48	42.65	46.65	62.14	47.25	53.68	71.12	57.20	63.41
JET-O	67.97	60.32	63.92	58.47	43.67	50.00	58.35	51.43	54.67	64.77	61.29	62.98
GTS-BERT	70.92	69.49	70.20	57.52	51.92	54.58	59.29	58.07	58.67	68.58	66.60	67.58
S ³ E ²	69.08	64.55	66.74	59.43	46.23	52.01	61.06	56.44	58.66	71.08	63.13	66.87
Dual-MRC	71.55	69.14	70.32	57.39	53.88	55.58	63.79	51.87	57.21	68.60	66.24	67.40
BMRC	-	-	70.01	-	-	57.83	-	-	58.74	-	-	67.49
Ours	73.20	71.31	72.23	64.27	53.65	58.48	69.38	60.12	64.24	71.24	68.80	69.99

TABLE 7. Experimental results (%) of the ASTE task on the ASTE-Data-V2 datasets. The baseline results with † are from [24]. Other baseline results are retrieved from the original papers. If the baseline to be compared has a Bert-based version, the BERT-based version is selected for the results. The best scores are marked in bold.

Models	14res			14lap			15res			16res		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Li-unified-R+†	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-two-stage†	43.24	63.66	51.46	37.38	50.38	42.87	48.07	57.51	52.32	49.96	64.24	54.21
JET-T	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	70.56	55.94	62.40	55.39	47.33	51.04	64.45	51.96	57.53	70.42	58.37	63.83
EIN	71.75	70.52	71.13	65.25	53.79	58.97	62.77	59.79	61.25	68.20	69.26	68.73
ASTE-RL	70.60	68.65	69.61	64.80	54.99	59.50	64.45	60.29	62.72	67.21	69.69	68.41
UniASTE	72.14	66.30	69.09	62.24	51.77	56.51	64.83	54.31	59.05	69.06	65.53	67.22
Ours	74.12	72.84	73.47	65.63	55.04	59.86	67.75	60.33	63.82	71.17	67.84	69.46

E. EXPERIMENTAL RESULTS

The main results of the ASTE task on the ASTE-Data-V1 datasets and ASTE-Data-V2 datasets are listed in Tables 6 and 7, respectively. Since the standard deviation is at most 0.39 and 0.42 in both datasets, so we omitted it. According to the results, the framework we proposed greatly exceeds all baseline results on both benchmark datasets, indicating that it has good robustness and comprehensive performance. And compared with the GTS-BERT framework [23], our framework achieves an apparent absolute F1-score increase of 2.03%, 3.9%, 5.57%, and 2.41% on 14res, 14lap, 15res, and 16res, respectively. The results show that the increase of the F1-score is mainly due to more significant precision, which indicates that the proposed framework has a higher prediction accuracy for the ASTE task than the baselines. In addition, the F1-score of the end-to-end approaches is generally higher than that of the pipeline approaches, which implies that establishing connections between subtasks can help improve the performance of the ASTE task. Furthermore, it is also found that the precision of the pipeline approaches is usually lower than their recall, while the precision scores of the end-to-end approaches and the MRC approaches are usually higher than their recall, which shows that the pipeline approaches are good at retrieving more triplets, while the end-to-end approaches and MRC approaches may pay more attention to the accuracy of triplet extraction.

V. EXTENSIVE EXPERIMENTS

In this section, the performance of the model is explored through extensive experiments. These experiments mainly

TABLE 8. F1-score (%) for ablation study on the ASTE task. The best scores are marked in bold.

Models	14res	14lap	15res	16res
Ours	72.23	58.48	64.24	69.99
w/o Label Name BERT encoder	70.99	57.08	62.33	68.15
w/o multi-task	71.54	57.95	61.33	68.39
w/o inference	71.51	57.85	61.14	69.14
Use six-tag	71.59	57.57	61.47	67.86

include ablation study, the extraction performance of the model on subtasks, the effect of terms length, and the effect of the auxiliary tasks hyperparameter on the experimental results. Finally, a case study is conducted to compare with other methods.

A. ABLATION STUDY

To estimate the effectiveness of the different modules in the proposed framework, an ablation study is conducted on the ASTE-Data-V1 datasets. The results of the ablation study are shown in Table 8.

w/o Label Name BERT encoder denotes removal of the Label Name BERT encoder from the framework. It can be seen that this simple feature information fusion can result in significant information gain; additionally, the F1-score drops the most (1.91%) on the 15res datasets and the least (1.24%) on the 14res datasets, indicating that the Label Name BERT encoder is suitable for datasets with a small amount of sample data, and has a greater impact on these datasets. In addition, compared to [32] using a graph convolutional

neural network (GCN) to learn features and improve the average F1-score by 1.19%, the Label Name BERT encoder uses a simpler method to improve the F1-score by 1.59%. *w/o* multi-task means that only a single task framework is used; that is, only the main task framework is employed to train the framework. The decline of the F1-score is also a good indication that the auxiliary task can make the framework focus on those influential representation features, and by exacting greater penalties to help the main task in undertaking better grid tagging. *w/o* inference means that the framework is not able to use potential constraints between word pairs to infer. It can be seen that without the inference constraints, the baseline [23] can still be exceeded, which indicates that the proposed dual-encoder is efficient. Using six-tag means that we use the six tags proposed in [23] to tag the grid. When the unified tags lose the relative position information and the related inference constraints of terms, the F1-score is decrease, implying that the prediction and the tagging abilities of the main task have also declined. We further elucidate the role of the ten-tag tagging approach through a case study in Section V-E. In conclusion, the results of the ablation study suggest that when any one module is removed from the framework, it leads to a large drop in the F1-score, which indicates that each module constructed in the framework is effective.

B. RESULTS ON SUBTASKS

To investigate the ability of the framework to extract different elements of triplets, the F1-score of our framework is compared with JET-T, JET-O, GTS, and UniASTE on three subtasks of ATE, OTE, and AOPE. The results are shown in Figure 4. According to the results, both JET-T and JET-O are limited by the inherent limitations of position awareness and are inaccurate enough for the extraction of the three subtasks. Although the comprehensive performance of UniASTE has improved, it is limited to only predicting boundaries and cannot sufficiently interact with the associations between aspect terms, opinion terms, and sentiment polarity. Additionally, GTS lacks the feature information of the word itself and places each word that constitutes the aspect term and opinion term in an equally important position, ignoring its relative position information, its score on the three subtasks is still not high. In contrast, our framework achieves the best results compared with the baseline results, which shows that the proposed framework is good at dealing with these subtasks, and can use the implicit associations and constraints to extract each element in the review sentences.

C. EFFECT OF TERMS LENGTH

To investigate the effect of terms length on triplet extraction, we count the res14 and lap14 testsets of the ASTE-Date-V2 datasets with the following settings and redivide them into three new sub-testsets: #1: Aspect terms and opinion terms in the triplets are composed of a single word; #2: There are aspect terms or opinion terms composed of two words in the triplets; #3+: There are aspect terms or opinion terms composed of three or more words in the triplets.

TABLE 9. F1-score of different evaluation modes on the ASTE task.

Models	14res			14lap		
	#1	#2	#3+	#1	#2	#3+
GTS-BERT	79.79	56.84	43.61	68.27	47.97	25.65
Ours	81.44	61.15	51.35	68.94	50.22	29.94
Δ	+1.65	+4.31	+7.74	+0.67	+2.25	+4.29

We compare the performance of our model with the previous work [23] on the above three settings and report the average F1-score of five times with different random seeds in Table 9. It can be seen from the table that no matter under which setting, the extraction performance of our model is always higher than that of [23], and as the length of the terms increases, the difference in model performance becomes increasingly larger, which indicates that our proposed inference strategies are effective, and good at solving the triplet extraction task according to the importance of words in different positions in multiple words.

D. EFFECT OF THE AUXILIARY TASK HYPERPARAMETER

In this section, four values of the hyperparameter α are set to 0.001, 0.01, 0.1, and 1, respectively, and the effect of the auxiliary task on the performance of the main task is studied by adjusting the value of the hyperparameter α . The results are shown in Figure 5.

Any value of the hyperparameter that is too small or too large will affect the performance of the main task. From 0.001 to 0.01, the polylines in both figures show an upward trend, which means that with the increasing influence of the auxiliary task on the main task, the F1-score gradually increases. At values greater than 0.01, it can be seen that except for the polyline 16res, the rest of the experimental results show a downward trend, indicating that the influence of the auxiliary task on the main task is negative. We infer that the reason why 16res is different from other datasets may be that the aspect terms or opinion terms composed of multiple words in the 16res datasets account for a small proportion. So, after inference, the prediction of the auxiliary task tends to be consistent with the prediction of the main task on the main diagonal, and still has a positive influence within this range. Finally, after a comprehensive analysis, the hyperparameter α of the auxiliary task is set to 0.01.

E. CASE STUDY

In order to further demonstrate the effectiveness of the proposed ten-tag tagging approach, the extraction results using the six-tag and the ten-tag tagging approaches based on five examples are compared. Of the five examples, the second and third are taken from the laptop domain, and the rest are taken from the restaurant domain. The results are shown in Table 10.

For the first example, the sentence is relatively simple, and the correct triplet can be accurately extracted using either six-tag or ten-tag tagging approach. For the second example and the third example, the aspect term “*delivery times*” and

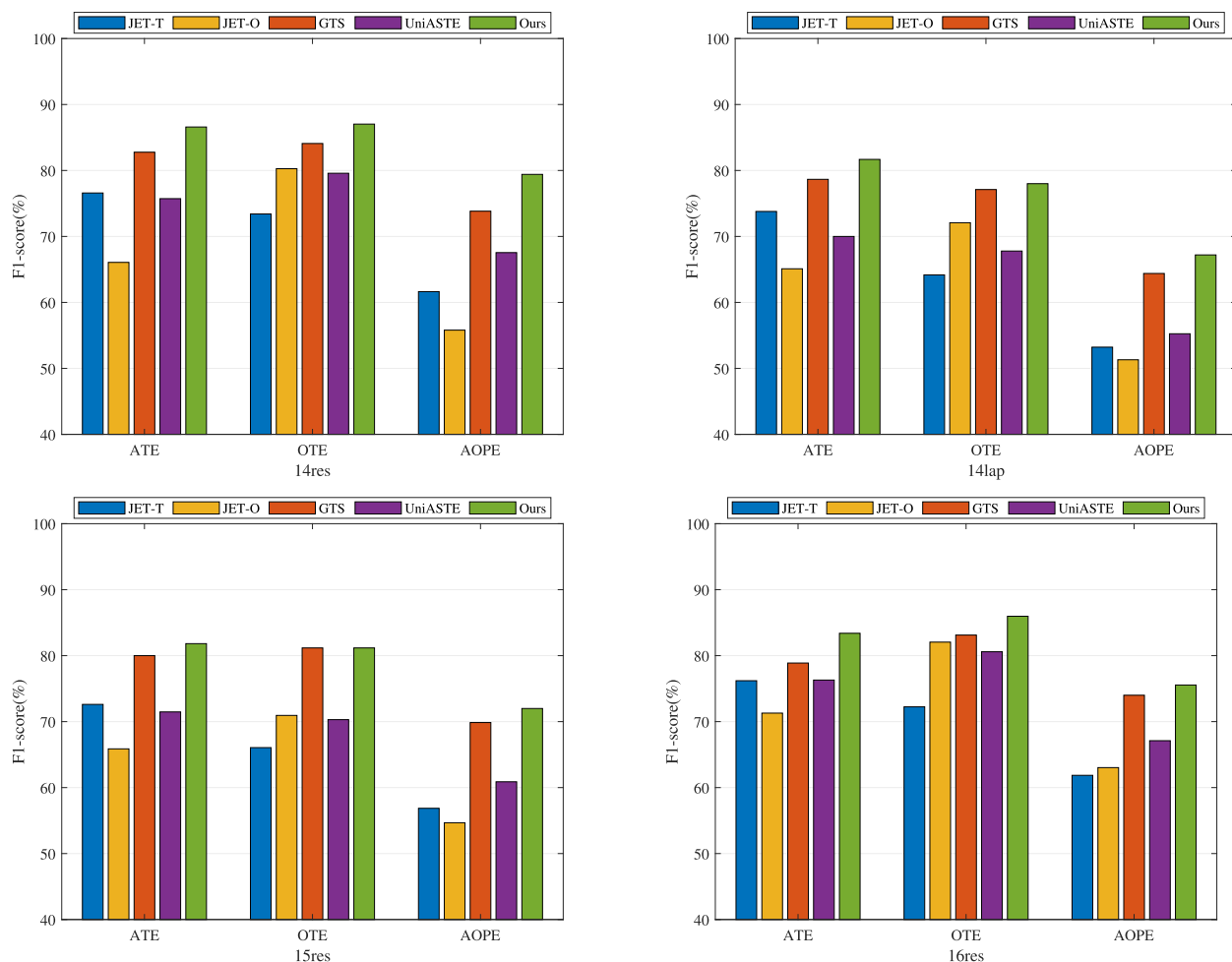


FIGURE 4. Performance on ATE, OTE, and AOE tasks.

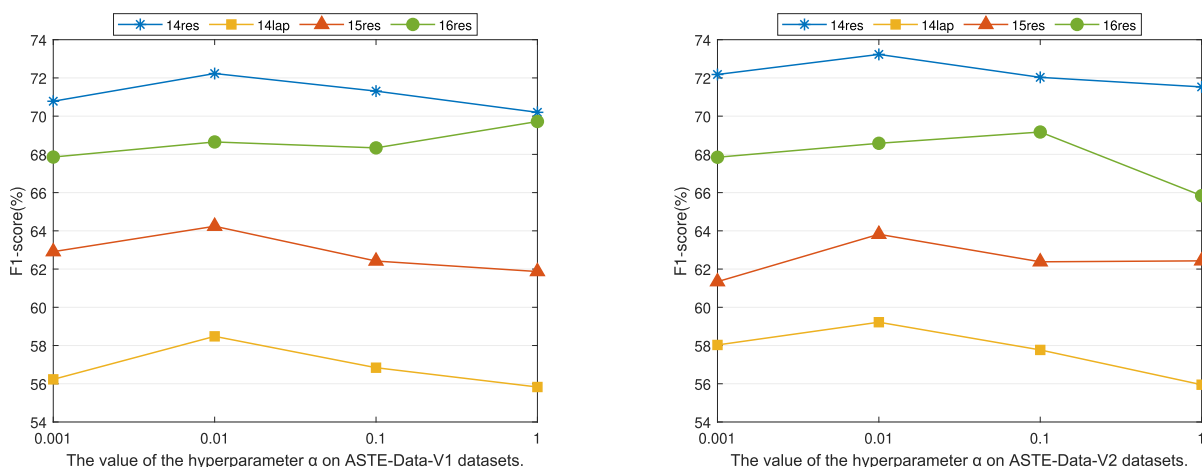


FIGURE 5. Effect of the auxiliary task hyperparameter α on the F1-score of the main task.

the opinion term “not fix” are composed of multiple words. It can be seen that the ten-tag tagging approach can correctly extract the corresponding triplet, while the six-tag tagging approach produces errors and only extracts “delivery” and

“not”, which proves that the feature enhancement method we used is effective, and our framework is good at using the enhanced features to search for complete aspect terms and opinion terms. For the fourth example, the triplets in it cannot

TABLE 10. Results of case study. Incorrect results are marked with X.

Examples	Ground Truth	Use six-tag	Use ten-tag
S1 : Did I mention that the coffee is outstanding?	(coffee,outstanding,positive)	(coffee,outstanding,positive)	(coffee,outstanding,positive)
S2 : I have to say they have one of the fastest delivery times in the city.	(delivery times,fastest,positive)	(delivery,fastest,positive)X	(delivery times,fastest,positive)
S3 : tech support would not fix the problem unless I bought your plan for \$ 150 plus.	(tech support,not fix,negative)	(tech support,not,negative)X	(tech support,not fix,negative)
S4 : The receiver was full of superlatives for the quality and performance.	(quality,superlatives,positive) (performance,superlatives,positive)	EmptyX	(receiver,superlatives,positive)X (quality,superlatives,positive) (performance,superlatives, positive)
S5 : It was pleasantly uncrowded, the service was delightful, the garden adorable , the food (from appetizers to entrees) was delectable.	(service,delightful,positive) (garden,adorable,positive) (food,delectable,positive) (appetizers,delectable,positive) (entrees,delectable,positive)	(service,delightful,positive) (garden,adorable,positive) (food,delectable,positive) (entrees,delectable,positive) missing one tripletX	(service,delightful,positive) (garden,adorable,positive) (food,delectable,positive) (appetizers,delectable,positive) (entrees,delectable,positive)

be identified by the six-tag tagging approach. Although our framework extracts all the correct triplets, there is an error due to the extraction of a surplus triplet (*receiver, superlatives, positive*). The reason for the error is that the multi-extracted aspect term “*receiver*” incorporates the noun part-of-speech feature and ignores the grammatical structure information, causing the framework to make erroneous predictions. For the fifth example sentence, the aspect terms and opinion terms are composed of a single word, but the sentence structure is very complex. There is a many-to-one situation between the aspect terms and the opinion terms. The six-tag tagging approach can only extract four triplets without one triplet (*appetizers, delectable, positive*). However, the ten-tag tagging approach can completely find all the triplets, indicating that our approach is good at coping with the complex many-to-one relationships between aspect terms and opinion terms.

VI. CONCLUSION

In this paper, a multi-task dual-encoder framework that incorporates semantic information is proposed to complete the ASTE task in an end-to-end fashion in one pass. The framework constructs two encoders: token BERT encoder and Label Name BERT encoder. The original token features are combined with word part-of-speech features and then multi-task joint training is performed. One of the multi-task is the word pair relationship tagging main task, and the other is the boundary prediction auxiliary task. At the same time, the different importance of words in different positions is considered. The beginning of the terms is the extraction center, and the last is the sentiment center. Based on this, two effective inference strategies and a decoding algorithm are constructed by adding implicit constraints. With the above-mentioned rich structure, extensive experiments are conducted on two benchmark datasets. The experimental results show that the proposed framework is simple and effective, outperforming all the compared baselines. For future work, we will consider using a span-based model with GCN to learn and integrate more semantic grammar information to identify aspect terms and opinion terms.

REFERENCES

- [1] M. Hu and B. Liu, “Mining and summarizing customer reviews,” in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2004, pp. 168–177.
- [2] T. T. Thet, J.-C. Na, and C. S. G. Khoo, “Aspect-based sentiment analysis of movie reviews on discussion boards,” *J. Inf. Sci.*, vol. 36, no. 6, pp. 823–848, 2010.
- [3] R. Obiedat, D. Al-Darras, E. Alzaghouli, and O. Harfoushi, “Arabic aspect-based sentiment analysis: A systematic literature review,” *IEEE Access*, vol. 9, pp. 152628–152645, 2021.
- [4] H. Xu, B. Liu, L. Shu, and P. S. Yu, “Double embeddings and CNN-based sequence labeling for aspect extraction,” in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics (Short Papers)*, vol. 2, 2018, pp. 592–598.
- [5] D. Ma, S. Li, F. Wu, X. Xie, and H. Wang, “Exploring sequence-to-sequence learning in aspect term extraction,” in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 3538–3547.
- [6] A. S. Fadel, M. E. Saleh, and O. A. Abulnaja, “Arabic aspect extraction based on stacked contextualized embedding with deep learning,” *IEEE Access*, vol. 10, pp. 30526–30535, 2022.
- [7] A. Pouran Ben Veyseh, N. Nouri, F. Dernoncourt, D. Dou, and T. H. Nguyen, “Introducing syntactic structures into target opinion word extraction with deep learning,” in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 8947–8956.
- [8] J. Jiang, A. Wang, and A. Aizawa, “Attention-based relational graph convolutional network for target-oriented opinion words extraction,” in *Proc. 16th Conf. Eur. Chapter Assoc. Comput. Linguistics, Main*, 2021, pp. 1986–1997.
- [9] Y. Wang, M. Huang, X. Zhu, and L. Zhao, “Attention-based LSTM for aspect-level sentiment classification,” in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 606–615.
- [10] C. R. Aydin and T. Gungor, “Combination of recursive and recurrent neural networks for aspect-based sentiment analysis using inter-aspect relations,” *IEEE Access*, vol. 8, pp. 77820–77832, 2020.
- [11] A. H. Sweidan, N. El-Bendary, and H. Al-Feel, “Sentence-level aspect-based sentiment analysis for classifying adverse drug reactions (ADRs) using hybrid ontology-XLNet transfer learning,” *IEEE Access*, vol. 9, pp. 90828–90846, 2021.
- [12] X. Li, X. Fu, G. Xu, Y. Yang, J. Wang, L. Jin, Q. Liu, and T. Xiang, “Enhancing BERT representation with context-aware embedding for aspect-based sentiment analysis,” *IEEE Access*, vol. 8, pp. 46868–46876, 2020.
- [13] K. Sattar, Q. Umer, D. G. Vasbieva, S. Chung, Z. Latif, and C. Lee, “A multi-layer network for aspect-based cross-lingual sentiment classification,” *IEEE Access*, vol. 9, pp. 133961–133973, 2021.
- [14] H. Shim, D. Lowet, S. Luca, and B. Vanrumste, “LETS: A label-efficient training scheme for aspect-based sentiment analysis by using a pre-trained language model,” *IEEE Access*, vol. 9, pp. 115563–115578, 2021.
- [15] X. Ye, Y. Xu, and M. Luo, “ALBERTC-CNN based aspect level sentiment analysis,” *IEEE Access*, vol. 9, pp. 94748–94755, 2021.

- [16] L. Xiao, Y. Xue, H. Wang, X. Hu, D. Gu, and Y. Zhu, "Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks," *Neurocomputing*, vol. 471, pp. 48–59, Jan. 2022.
- [17] S. Wu, H. Fei, Y. Ren, B. Li, F. Li, and D. Ji, "High-order pair-wise aspect and opinion terms extraction with edge-enhanced syntactic graph convolution," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 29, pp. 2396–2406, 2021.
- [18] S. Chen, J. Liu, Y. Wang, W. Zhang, and Z. Chi, "Synchronous double-channel recurrent network for aspect-opinion pair extraction," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 6515–6524.
- [19] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An interactive multi-task learning network for end-to-end aspect-based sentiment analysis," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 504–515.
- [20] Z. Chen and T. Qian, "Retrieve-and-Edit domain adaptation for End2End aspect based sentiment analysis," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 659–672, 2022.
- [21] H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, and L. Si, "Knowing what, how and why: A near complete solution for aspect-based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 5, Apr. 2020, pp. 8600–8607.
- [22] H. Yan, J. Dai, T. Ji, X. Qiu, and Z. Zhang, "A unified generative framework for aspect-based sentiment analysis," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process. (Long Papers)*, vol. 1, 2021, pp. 2416–2429.
- [23] Z. Wu, C. Ying, F. Zhao, Z. Fan, X. Dai, and R. Xia, "Grid tagging scheme for aspect-oriented fine-grained opinion extraction," in *Proc. Findings Assoc. Comput. Linguistics, (EMNLP)*, 2020, pp. 2576–2585.
- [24] L. Xu, H. Li, W. Lu, and L. Bing, "Position-aware tagging for aspect sentiment triplet extraction," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2020, pp. 2339–2349.
- [25] L. Xu, Y. K. Chia, and L. Bing, "Learning span-level interactions for aspect sentiment triplet extraction," in *Proc. 59th Annu. Meeting Assoc. Comput. Linguistics 11th Int. Joint Conf. Natural Lang. Process. (Long Papers)*, vol. 1, 2021, pp. 4755–4766.
- [26] P. Wang, T. Liu, D. Dai, R. Xu, B. Chang, and Z. Sui, "Explicit interaction network for aspect sentiment triplet extraction," 2021, *arXiv:2106.11148*.
- [27] Y. Mao, Y. Shen, C. Yu, and L. Cai, "A joint training dual-MRC framework for aspect based sentiment analysis," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 15, 2021, pp. 13543–13551.
- [28] Y. Li, Y. Lin, Y. Lin, L. Chang, and H. Zhang, "A span-sharing joint extraction framework for harvesting aspect sentiment triplets," *Knowl.-Based Syst.*, vol. 242, Apr. 2022, Art. no. 108366.
- [29] S. Chen, Y. Wang, J. Liu, and Y. Wang, "Bidirectional machine reading comprehension for aspect sentiment triplet extraction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 14, 2021, pp. 12666–12674.
- [30] L. Huang, P. Wang, S. Li, T. Liu, X. Zhang, Z. Cheng, D. Yin, and H. Wang, "First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction," 2021, *arXiv:2102.08549*.
- [31] F. Chen, Z. Yang, and Y. Huang, "A multi-task learning framework for end-to-end aspect sentiment triplet extraction," *Neurocomputing*, vol. 479, pp. 12–21, Mar. 2022.
- [32] Z. Chen, H. Huang, B. Liu, X. Shi, and H. Jin, "Semantic and syntactic enhanced aspect sentiment triplet extraction," in *Proc. Findings Assoc. Comput. Linguistics, (ACL-IJCNLP)*, 2021, pp. 1474–1483.
- [33] S. Yu Bai Jian, T. Nayak, N. Majumder, and S. Poria, "Aspect sentiment triplet extraction using reinforcement learning," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2021, pp. 0607–3603.
- [34] X. Li, L. Bing, P. Li, and W. Lam, "A unified model for opinion target extraction and target sentiment prediction," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, Jul. 2019, pp. 6714–6721.
- [35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2019, pp. 4171–4186.
- [36] J. Ma, M. Ballesteros, S. Doss, R. Anubhai, S. Mallya, Y. Al-Onaizan, and D. Roth, "Label semantics for few shot named entity recognition," 2022, *arXiv:2203.08985*.
- [37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6000–6010.
- [38] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [39] P. Qi, Y. Zhang, Y. Zhang, J. Bolton, and C. D. Manning, "Stanza: A Python natural language processing toolkit for many human languages," in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics, Syst. Demonstrations*, 2020, pp. 101–108.
- [40] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, 2014, pp. 27–35.
- [41] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, and I. Androutsopoulos, "SemEval-2015 task 12: Aspect based sentiment analysis," in *Proc. 9th Int. Workshop Semantic Eval. (SemEval)*, 2015, pp. 486–495.
- [42] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. D. Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryigit, "SemEval-2016 task 5: Aspect based sentiment analysis," in *Proc. 10th Int. Workshop Semantic Eval. (SemEval)*, 2016, pp. 19–30.
- [43] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in Adam," 2017, *arXiv:1711.05101*.
- [44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, Jun. 2014.



HAI HUAN (Member, IEEE) received the B.E. and M.E. degrees from Nanjing Normal University, China, in 2004, and the Ph.D. degree from Niigata University, in 2008. He is currently an Associate Professor with the Nanjing University of Information Science and Technology. His current research interests include computer vision, natural language processing, and super-resolution reconstruction.



ZICHEN HE received the B.Eng. degree in computer science from Yangtze University, China, in 2021. He is currently pursuing the M.Sc. degree in electronic information with the Nanjing University of Information Science and Technology. His research interests include aspect based sentiment analysis and machine learning.



YAQIN XIE received the Ph.D. degree in communication and information systems from Southeast University, in 2011. She joined the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology. From February 2015 to February 2016, she was supported by the National Study Fund to visit King's College London, U.K., as a Visiting Scholar. Her current research interests include indoor localization, satellite navigation, routing planning, and artificial intelligence.



ZELIN GUO received the B.Eng. degree in electronic information engineering from Wuxi University, China, in 2020. He is currently pursuing the M.Sc. degree in electronic and communication engineering with the Nanjing University of Information Science and Technology. His research interests include text classification and machine learning.

...