Hybrid of Spans and Table-Filling for Aspect-Level Sentiment Triplet Extraction

Minghua Nuo[1,2,3] * , **Chaofan Guo**[1]
[1] College of Computer Science, Inner Mongolia University, Hohhot, Inner Mongolia 010021, China
[2] National & Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, Hohhot, Inner Mongolia 010021, China
[3] Inner Mongolia Key Laboratory of Mongolian Information Processing Technology, Hohhot, Inner Mongolia 010021, China
nuominghua@163.com, 32109130@mail.imu.edu.cn

Abstract

Aspect Sentiment Triplet Extraction (ASTE) has become an emerging task in sentiment analysis research. Recently, researchers have proposed different tagging schemes, containing tagging of words, tagging of word-pairs, and tagging of spans. However, the first two of these methods are often insufficient for the identification of multi-word terms, while the span tagging can label the entire phrase span, but it lacks the interactive information between words. In this paper, we propose Hybrid of Spans and Table-filling (S&T) model which combining span with table-filling. Specifically, S&T model achieve full fusion of syntactic and contextual features through cross-attention and generate the structures of word-pair table through Biaffine. Then, our model converts it to a span table by computing semantic distance based on syntactic dependency tree, which can enrich each unit of span table with semantic and interactive information. Meanwhile, the initial sentence features are constructed as simple phrase tables to enhance textual information of the phrase itself. In decoding, we define 8 types of labels for identifying three dimensions including aspect, opinion, and sentiment. Finally, the extensive experiments on widely-used dataset show S&T model achieves competitive results in ASTE task, the results certify the effectiveness and robustness of our S&T model.

Keywords: Aspect Sentiment Triplet Extraction, Span, Table-Filling, Cross-Attention

1. Introduction

Sentiment Analysis (SA) (Zhao et al., 2010) task is an important research direction in the field of Natural Language Processing (NLP), which aims to analyze and reason about subjective texts with emotions. Sentiment analysis can be divided into coarse-grained sentiment analysis and fine-grained sentiment analysis. Coarse-grained sentiment analysis includes sentiment analysis at document or sentence level, and fine-grained sentiment analysis refers to attribute level sentiment analysis. Coarse-grained sentiment analysis can only perform sentiment analysis on the whole, and cannot satisfy people's real needs in reality, which led the birth of fine-grained sentiment analysis. Among them, ASTE task can deal with fine-grained sentiment analysis problems more comprehensively. As shown in Figure 1, the ASTE task aims to extract (aspect, opinion, sentiment polarity) triplets such as (hot dogs, top notch, POS) and (coffee, average, NEG).

The hot dogs are top notch but average coffee .

Aspects: hot dogs; coffee

Opinions: top notch; average

Triplets: (hot dogs, top notch, positive) ;
(coffee, average, negative)

Figure 1: An example of an ASTE task. Aspect items are highlighted in orange and opinion items in blue.

Table-filling method and span method are two commonly used model paradigms for ASTE task. Table-filling can fully calculate the word-pair information, and its end-to-end operation can avoid the error propagation caused by pipeline. In addition, table-filling fits maps it easier to calculate relationships between words, but when aspect or opinion is a multilingual term, the table-filling goes to determine the term boundaries, and term information refers to the entire term, not just the beginning and the end. In addition, the phenomenon of multiple sentiments may occur when decoding multi-word term sentiment, and the solution is heuristic rules, which will be affected by subjective. The span method judges the phrase as a whole, which is completer and more sufficient for multi-word information. Meanwhile, the result of sentiment polarity analysis is more objective. However, it is limited by the span size and the lack of information interaction within the phrase.

*Corresponding author.

LREC-COLING 2024, pages 8464–8473
20-25 May, 2024.

In this paper, we propose a *Hybrid of Spans and Table-filling* (S&T) model, that combines table-filling and the span to solve the ASTE task. Specifically, this paper uses an interactive mechanism to fully integrate the semantic information and context information. After that, Biaffine is used to calculate the interaction between words and generate the table structure, and then add semantic information to the table. Finally, S&T model converts the table structure into a span table which is used for label prediction after adding simple phrase information. We conduct experiments on public datasets to compare and analyze the results of multiple benchmark models. Then, ablation experiments were designed to explore the effects of each module. The results demonstrate the effectiveness of S&T model. In summary, the main work of this paper is as follows:

- Introducing information-enhanced cross-attention, using direction-guided graph convolutional network (D-GCN) to capture syntactic features, and enhance the fusion of context and syntactic feature. In addition, the Biaffine is used to locate aspects and opinions more accurately, thus improving the accuracy of extraction.

- In the observation, it is found that related words are more closely in semantic distance then relative positional distance. Therefore, we believe that the introduction of semantic distance can enhance the correlation information between aspects and corresponding opinions, and play a role in filtering noise.

- The table-filling enhances the computation of inter-word information, but lacks the whole information of multi-word terms. The span approach performs better for the whole calculation, but the size of the span is difficult to set. We proposes to combine the span and table-filling, construct span in the form of a table, and calculate the label based on the span, so as to combine the advantages of both.

- The S&T model is compared with the mainstream ASTE task model on common datasets. The experimental results show that the S&T model has improved to some extent.

2. Related Work

The ASTE task was first proposed by Peng et al. (2019) and a pipeline model was designed. Then Zhang et al. (2020), Xu et al. (2020), Yu et al. (2021) and Mukherjee et al. (2023) separately added multi-task learning, location bias, hierarchical reinforcement learning and contrast learning to improve the performance. However, the pipeline approach suffers from error propagation and a lack of information interaction between terms.

In order to enhance the computation of inter-word information and reduce error propagation, Wu et al. (2020) proposed GTS, which defined ASTE task as an end-to-end table labeling task. This method is also known as table-filling. Later, on the basis of GTS, Chen et al. (2021b) proposed using GCN to encode syntactic information and concatenate it to table; Hu et al. (2022) added knowledge graphs and the attention module. Chen et al. (2022a) designed a ten-label method to enhance the recognition of boundary information and improve the model performance. Liu et al. (2023) cast ASTE as a boundary words relation classification. Yuan et al. (2023) proposes a syntax-aware transformer. Table-filling can be convenient to realize the calculation of information exchange between words, but when there is a multiple word, the overall information will be ignored.

For utilize the complete information of multiple items, researchers proposed span calculation. Xu et al. (2021) proposed a span interaction model. Li et al. (2022) suggest that aspects and opinions can share the span. Later, Chen et al. (2022c) and Zhang et al. (2022b) proposed multi-layer attention network and multi-layer GRU respectively in order to better deal with multi-term and overlapping problems. Chen et al. (2022b) proposed decoding from two directions. Wang et al. (2023) added more additional information when constructing the span. The span approach can make use of the overall information of the term, but the basis of the span approach is still the pipeline approach.

Seq2Seq generate(Sutskever et al., 2014) and machine reading comprehension(MRC)(Liu et al., 2019) structures are also used for the ASTE task. Zhang et al. (2021) and Yan et al. (2021) turned the ASTE task into a generative problem, respectively using T5(Raffel et al., 2019) and BART(Chipman et al., 2008). Fei et al. (2021) proposed a non-autoregressive decoding method. Mao et al. (2021) designed two MRC to achieve triplet extraction. Chen et al. (2021a) designed a BMRC to improve accuracy. Liu et al. (2022) resume span matching, adding word segmentation and classifier to BMRC to improve model robustness. Besides, Yu et al. (2023) proposes a retrieval-based ASTE approach.

3. Methodology

The framework for S&T model is showed in Figure 2, and it consists of five parts: sentence feature representation, iterative fusion of features, calculation of span table and decoding inference. The detailed descriptions of each part are given in the following subsections.

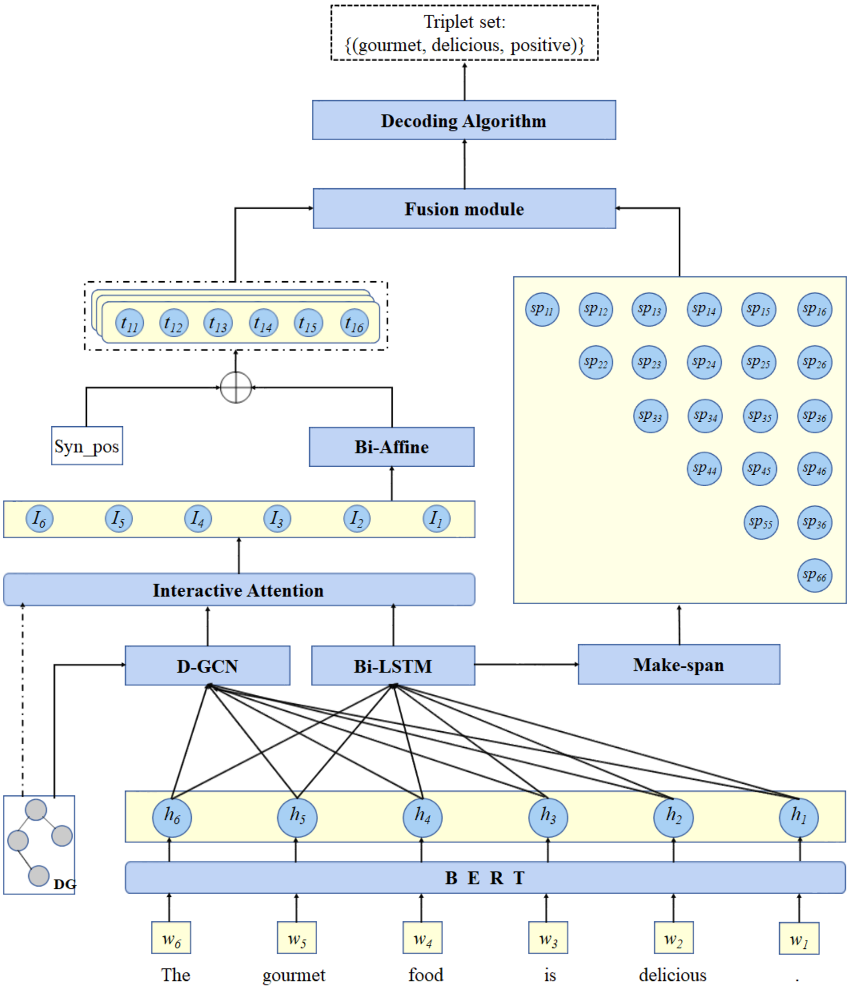

Figure 2: Overall architecture of the S&T model

3.1. Task Definition

Given an input sentence $X = \{w_1, w_2, \ldots, w_n\}$ with n words, the ASTE task aims to obtain a set of triples $T = \{(a, o, s)_m\}_{m=1}^{|T|}$, where a and o respectively denote aspect and opinion item. The sentiment polarity s belongs to the given sentiment label set S = {POS, NEU, NEG}. S includes three sentiment polarities: positive, neutral and negative, and $|T|$ is the total number of triples which insist in sentence X.

3.2. Sentence Feature Representation

We use BERT (Devlin et al., 2019) as sentence encoder to extract the hidden contextual representation. Given a sentence $X = \{w_1, w_2, \ldots, w_n\}$ with n words, the encoding layer outputs a sequence of hidden representations $H = \{h_1, h_2, \ldots, h_n\}$. The specific processing BERT can be expressed as:

$$H = BERT(T([CLS] + w_1, w_2, \ldots, w_n + [SEP])) \tag{1}$$

where T(-) represents the splitter used by BERT, BERT(-) represents the encoding model, and "[CLS]" and "[SEP]" are sentence identifiers added by BERT input as needed, the output of the last layer of the transformer can represent initial embedding of the sentence.

Through BERT processing, the token length may be inconsistent with the length of the original sentence, resulting in the context feature dimension are inconsistent with the original sentence. So, the average method shown in Eq.(2) is used to align the sub words and summarize the semantic representation of the words.

$$\hat{h}_k = \frac{1}{|T(X)|} \sum_{m \in T(X)} h_m \tag{2}$$

where || returns its length.

3.3. Fusion of Features

After word embeddings are obtained, BiLSTM is used to learn contextual features, D-GCN is used to learn syntactic features, and the two features are iteratively fused in the interactive module where mainly uses cross-attention mechanism.

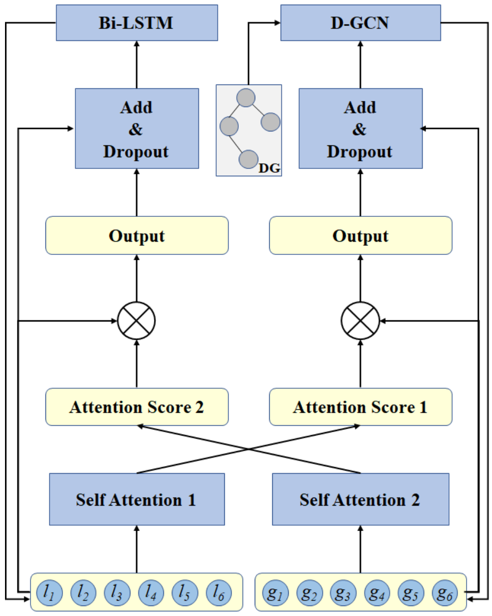

Figure 3: The structure of interactive module

The iterative module is shown in Figure 3.

$$\hat{h}_i^c = [\overrightarrow{h_{LSTM}}(\hat{h}_i) \bigoplus \overleftarrow{h_{LSTM}}(\hat{h}_i)] \quad (3)$$

$$h_i^c = Linear(\hat{h}_i^c) \quad (4)$$

where $\hat{h}_i^c \in R^{2d_l}$ is the output of the BiLSTM, d_l is the output dimension of unidirectional LSTM, the output of BiLSTM is projected to a low-dimensional space using a fully connected layer, and $h^c \in R^{d_l}$ is the context representation after the linear layer.

In order to enhance the ability of syntactic information representation of input sentence, we use D-GCN to compute syntactic dependencies between words.The initial input is the output of BERT and the adjacency matrix of the sentence. Then the l^{th}-layer D-GCN output is calculate as follows:

$$g_i^{(l)} = ReLU(\sum_{j=1}^{n} p_{i,j}^{(l)}(W_{dir}^{(l)} \cdot g_i^{(l-1)} + b^{(l)})) \quad (5)$$

where $W_{dir}^{(l)}$ and $p_{i,j}^{(l)}$ are learnable parameters. $W_{dir}^{(l)}$ encodes the positional relationship between word-pairs (x_i, x_j) in three forms: $W_{left}^{(l)}$, $W_{right}^{(l)}$ and $W_{self}^{(l)}$, respectively represent the positional information between x_i, x_j and the information of x_i itself. In addition, in the D-GCN, the weights of words are expressed $p_{i,j}^{(l)}$. The $p_{i,j}$ is calculated as follows:

$$p_{i,j}^{(l)} = \frac{a_{i,j} \cdot exp(h_i^{(l-1)} \cdot h_j^{(l-1)})}{\sum_{j=1}^{n} a_{i,j} \cdot exp(h_i^{(l-1)} \cdot h_j^{(l-1)})} \quad (6)$$

where $a_{i,j}$ comes from the adjacency matrix A of the sentence, if there is an edge between x_i and x_j, the information between word-pairs can be calculated in Eq.(6).

In this paper, the interactive module contains two sub-attention layers, which are located after the BiLSTM and D-GCN. The computational procedure of the attention mechanism of our interactive module is shown from Eq.(7) to (12). In particular, the BiLSTM and D-GCN here have the same structure as the corresponding network before the interactive module.

$$\alpha_t^c = softmax(\frac{S_t^c \cdot S_t^{cT}}{\sqrt{d_l}}) \quad (7)$$

$$\alpha_t^g = softmax(\frac{S_t^g \cdot S_t^{gT}}{\sqrt{d_g}}) \quad (8)$$

$$S_t^{c'} = \alpha_t^g \cdot S_t^c \quad (9)$$

$$S_t^{g'} = \alpha_t^c \cdot S_t^g \quad (10)$$

$$S_{t+1}^c = BiLSTM(Dropout(S_t^{c'}) + S_t^c) \quad (11)$$

$$S_{t+1}^g = DGCN(Dropout(S_t^{g'}) + S_t^g) \quad (12)$$

where $S_t^c \in R^{n*d_l}$ and $S_t^g \in R^{n*d_g}$ are the outputs of BiLSTM and D-GCN after the t^{th} iteration, and $S_0^c = H^c$, $S_0^g = G$. Finally the output $S^c \in R^{n*d_l}$ of BiLSTM is chosen as the output of the interaction module.

3.4. Calculation of Span Table

Biaffine (Yu et al., 2020) has been shown to be sensitive to boundary information recognition. In S&T model, we firstly use Biaffine to calculate table, where each feature tensor represents a word-to-word relational feature. The calculation process is given in Eq.(13) and (14):

$$b_{i,j} = s_i^{c^T} W_1 h_j^c + W_2(h_i^c \bigoplus h_j^c) + b \quad (13)$$

$$t_{i,j,k}^b = \frac{exp\ (t_{i,j,k}^b)}{\sum_{l=1}^{d_b} exp((t_{i,j,l}^b)}) \quad (14)$$

Where $t_{i,j}^b \in R^{1*d_b}$ is the output of Biaffine, representing the modeling of the relations between word pairs (x_i, x_j), d_b is the size of Biaffine hidden dimension, W_1, W_2 and b are trainable parameters in the model, $\bigoplus$ represents the concat operation.

In our exploration, we find that syntactic distance can reflect the relation between words better than relative positional distance. Thus, we calculate the syntactic distance between word pairs and learn the corresponding embedding representation T^{dep}. By connecting the Biaffine representation T^b with the syntactic distance embedding T^{dep} obtained by StanfcoreNLP, the word pair table $t_{i,j} \in R^{1*d_b}$ is obtained, which carries the syntactic dependency and interaction information between words. Finally, Eq.(16) is used to convert

#	label	meaning
Aspect related	N	This phrase is not an aspect term.
	A	This phrase is an aspect term.
Opinion related	N	This phrase is not an opinion term.
	O	This phrase is an opinion term.
Sentiment related	N	Not constituting an aspect-opinion pair.
	POS	Two phrases constituting a positive sentiment.
	NEU	Two phrases constituting a neutral sentiment.
	NEG	Two phrases constituting a negative sentiment.

Table 1: Defined 3D labels and their meaning.

	The	gourmet	food	is	delicious	but	the	service	is	terrible
The	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N
gourmet		N-N-N	A-N-N	N-N-N	N-N-POS	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N
food			N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N
is				N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N
delicious					N-O-N	N-N-N	N-N-N	N-N-N	N-N-N	N-N-N
but						N-N-N	N-N-N	N-N-N	N-N-N	N-N-N
the							N-N-N	N-N-N	N-N-N	N-N-N
service								A-N-N	N-N-N	N-N-NEG
is									N-N-N	N-N-N
terrible										N-O-N

Figure 4: Span table labeling results for the ASTE task. The meanings of labels are refer to table 1.

the word pair table to the span table representation.

$$t_{i,j} = t^b_{i,j} \bigoplus t^{dep}_{i,j} \tag{15}$$

$$sp^t_{i,j} = \frac{1}{j-i+1} \sum_{k=i}^{j} t^c_{i,j} \tag{16}$$

To enhance the feature information of the multi-word phrase itself, we construct a simple phrase span table structure by tensor transformation and concat the table in Eq.(17) and table in Eq.(18).

$$sp^w_{i,j} = \frac{1}{j-i+1} \sum_{k=i}^{j} h^c_i \tag{17}$$

$$SP = SP^w \bigoplus SP^t \tag{18}$$

After obtaining the final span table representation, it is fed to the softmax with a fully connected layer to generate the decision space $p \in R^{d_p}$:

$$p = softmax(W_p sp + b_p) \tag{19}$$

where d_p is the number of labels and W_p and b_p are trainable weights and biases. Cross entropy loss function is used in the training process:

$$L = -\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k \in C} I(y_{i,j} = k) log\ (t_{i,j|k}) \tag{20}$$

where $y_{i,j}$ is the real label, $t_{i,j}$ is the predicted label distribution, *I*(-) is the indicator function, and C is the label set.

3.5. Label Definition and Decoding Inference

Considering that a phrase may play multiple roles simultaneously, three-dimensional composite labels are used in this paper, and the label name, definition and meaning are shown in Table 1. Label set consists of {N, A}-{N, O}-{N, POS, NEU, NEG}, where $role^a \in${N, A} is used to identify aspect, $role^o \in${N, O} is used to identify opinion, and $role^s \in${ N, POS, NEU, NEG} is used to identify aspect-opinion pairs is effective and judge its polarity.

Figure 4 shows the labeling results of an example sentence "The gourmet food is delicious but the service is terrible." The label for each cell consists of a three-dimensional label. Specifically, the label for "*gourmet food*" is "A-N-N", which means that it is an aspect, not an opinion, and indicate that it does not contain valid aspect-opinion pair.

The decoding algorithm in this paper adopts the greedy algorithm (Liang et al., 2022), where "greedy" refers to the word item with the largest length, and the decoding details are shown in Algorithm 1. Firstly, we traverse the upper triangle matrix to get all possible aspects, opinions and sentiment fragments, and then extract the sentiment triples based on the two possible cases in the sentiment fragments.

(1) Aspects appear before the opinions (lines 7 to 14). The beginning index of the sentiment segment is also the beginning index of the aspect in the segment, and the end index of the sentiment segment is also the end index of the opinion in the segment. If there are multiple candidates for a role at same time, the longest phrase is used to construct a valid triplet. When marking in three dimensions, the sentiment fragment itself may also be a candidate term, considered as an aspect or point of view if and only if there are no other candidates.

(2) Opinion appears before the aspect (lines 16 to 23). The beginning index of the sentiment segment is also the beginning index of the opinion in the segment, and the end index of the sentiment segment is also the end index of the aspect in the segment. The extraction process is the same as in case (1), so it will not be repeated here.

4. Experiment and Analysis

4.1. Datasets

In this paper, the ASTE-Data-V2(D2) dataset compiled and published by Xu et al. (2020) is used, which is from the SemEval ABSA challenges. The

Algorithm 1 Greedy algorithm based decoding of span level table-filling ASTE task

Input: The result P of the prediction of a sentence X of length n, where $P_{i,j}$ denotes the label prediction result span $SP_{i,j}$, and can be split 3D into $role^{a}_{i,j}$, $role^{o}_{i,j}$ and $role^{s}_{i,j}$, where a denotes aspect item relevance, o denotes opinion item relevance, and s denotes sentiment relevance.
Output: Given a set T of sentiment triples in a sentence X

```
Initialize D={}, A={}, O={}, T={}
A = {(i.j)|role^a_{i,j} = A, 0≤i≤n, i≤j≤n}
O = {(i.j)|role^o_{i,j} = O, 0≤i≤n, i≤j≤n}
D = {(i.j, role^s_{i,j})|role^s_{i,j}≠N, 0≤i≤n, i≤j≤n}
for each (i, j, s) in D do
  //CASE 1: aspect item before opinion item
  CA = {k|i≤k≤j, (i, k)∈A}
  CO = {l|i≤l≤j, (l, j)∈O}
  if CA≠∅ and CO≠∅ then
    remove j from CA when |CA| > 1
    remove i from CO when |CO| > 1
    k = max(CA), l = min(CO)
    T ← T ⋃(SP_{i,k}, SP_{l,j}, s)
  end if
  //CASE 2: opinion item before aspect item
  CO = {k|i≤k≤j, (i, k)∈O}
  CA = {l|i≤l≤j, (l, j)∈A}
  if CO ≠ ∅ and CA ≠ ∅ then
    remove j from CO when |CO| > 1
    remove i from CA when |CA| > 1
    k = max(CO), l = min(CA)
    T←T⋃(SP_{i,j}, SP_{l,k}, s)
  end if
end for
return T
```

dataset contains three datasets in the restaurant domain represented by res and one in the electronics domain represented by lap. The statistics are shown in Table 2.

Dataset		14res		14lap		15res		16res	
		#S	#T	#S	#T	#S	#T	#S	#T
D2	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

Table 2: Statistics of D2 dataset. #S represents the number of sentences in the dataset, and #T represents the number of aspect items.

4.2. Experimental Setup

In experiment, BERT-base-uncased model (Devlin et al., 2019) was used as the sentence encoder, and AdamW optimizer was selected. The sentence syntax information was obtained from StanfordCoreNLP. The settings of some experimental parameter are shown in table 3.

experimental parameter	value
Batch size	16
Max sequence length	102
Num of epoch	100
BERT learning rate	$2 * 10^{-5}$
AdamW learning rate	10^{-3}
Embedding dropout	0.5
Interactive dropout	0.3
BERT feature dimensions	768
LSTM feature dimension	300
Biaffine feature dimension	200
Grammatical feature dimension	50

Table 3: Partial experimental parameter settings

We implement three metrics to evaluate our proposed model: Precision rate(P), Recall rate(R) and F1 score. The extracted triplet is judged correct if and only if the aspect, opinion, and sentiment polarity are all correct. We select the best model based on the F1 score on development set. The results of the experiment are reported as the average of five runs using random seeds.

4.3. Baselines

To verify the validity of our S&T mode, the following baseline models were selected for performance comparison.

1) Sequence labeling methods:

JET(Xu et al., 2020) uses position-aware tagging scheme to capture word pair information and extract sentiment triples in combination.

EIN(Wang et al., 2021) uses two encoders for bidirectional explicit interaction.

FTTOP(Huang et al., 2021) achieves sentiment triplet extraction by inserting special markers generated in a stage into the sentence.

CONTRASTE-Base(Mukherjee et al., 2023) uses contrast learning to improve the performance of ASTE.

2) Table-filling methods:

GTS-BERT(Wu et al., 2020) aims to solve ASTE task in end-to-end mode with a unified table.

EMC-GCN(Chen et al., 2022a) uses multi-channel graphs to fuse multiple language features.

STAGE(Liang et al., 2022) simply spliceds the start word, the end word, and the overall feature.

BDTF(Zhang et al., 2022a) transforms the ASTE task into a region detection and classification task.

HIM(Liu et al., 2023) cast ASTE as a boundary-words relation classification problem.

SA-Transformer(Yuan et al., 2023) proposes a syntax-aware transformer that can encode dependency type information into both edge and word representations.

3) Span based methods:

Model	Res14			Lap14			Res15			Res16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
JET§	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
FTTOP∗	63.59	**73.44**	68.16	57.84	59.33	58.58	54.53	63.60	58.59	63.57	71.98	67.52
CONTRASTE-base∗	72.40	73.20	72.80	63.90	59.10	61.40	62.60	67.20	64.80	72.10	73.90	73.00
GTS-BERT§	68.14	68.77	68.45	58.62	52.35	55.29	62.37	59.71	60.98	66.16	68.81	67.44
EMC-GCN§	70.37	72.84	71.58	59.61	56.30	57.90	60.45	62.72	61.55	63.43	72.63	67.69
STAGE§	**79.54**	68.47	73.58	**71.48**	53.97	61.49	**72.05**	58.23	64.37	**78.38**	69.10	73.45
BDTF∗	75.53	73.24	74.35	68.94	55.97	61.74	68.76	63.71	66.12	71.44	**73.13**	72.27
HIM∗	76.99	70.46	73.57	65.99	56.05	60.59	69.65	63.23	66.25	73.11	71.05	72.06
SA-Transformer∗	70.76	65.85	68.22	61.28	48.98	54.44	62.82	58.31	60.48	72.01	62.87	67.13
Span-dual#	72.12	73.14	72.62	62.36	**60.37**	61.35	64.27	60.73	62.45	68.74	71.79	70.23
SSJE∗	73.12	71.43	72.26	67.43	54.71	60.41	63.94	**66.17**	65.05	70.82	72.00	71.38
ES-ASTE∗	71.01	68.34	69.67	66.43	52.31	59.37	60.26	65.02	62.63	65.18	64.46	64.82
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
Dual-MRC#	71.55	69.14	70.32	57.39	53.88	55.58	63.78	51.87	57.21	68.60	66.24	67.40
BART∗	66.52	64.99	65.25	61.41	56.19	58.69	59.14	59.38	59.26	66.60	68.68	67.62
RLI∗	77.46	71.97	74.34	63.32	57.43	60.96	60.08	70.66	65.41	70.50	**74.28**	72.34
S&T(ours)	76.85	72.19	**74.44**	70.26	56.86	**62.74**	71.91	61.48	**66.26**	76.72	70.50	**73.47**

Table 4: Experimental results on dataset D2. **§** indicates that the experimental results are from the paper (Liang et al., 2022), **#** indicates that the results are from (Chen et al., 2022c), and $*$ indicates that the results are from the original author's paper. Bolding denotes the best results and underlining denotes the second best results

Model	Res14			Lap14			Res15			Res16		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
S&T	76.85	72.19	74.44	70.26	56.86	62.74	71.91	61.48	66.26	76.72	70.50	73.47
S&T$_{simple_word}$	75.64	71.37	73.42	68.69	56.83	62.07	71.19	61.69	66.07	76.96	67.96	72.12
S&T$_{only_pair}$	76.79	70.68	73.57	67.34	57.41	61.87	70.66	60.70	65.25	76.59	69.50	72.84
S&T$_{word_no_span}$	74.22	70.32	72.21	66.58	56.77	61.26	68.96	61.69	65.10	75.59	69.26	72.25

Table 5: Results of ablation experimental research. The displayed scores are precision rate(P), recall rate(R) and F1 score(F1) for triplets extraction on D2 dataset.

Span-dual(Chen et al., 2022c) improves items extraction with two transformer-based multi-head attention decoders.

SSJE(Li et al., 2022) proposes a span sharing joint extraction model, which argues that span can be both an aspect and an opinion term.

ES-ASTE(Wang et al., 2023) uses GCN to introduce syntactic information and part-of-speech combinations.

4) Other end-to-end methods

UniASTE(Zhang et al., 2020) introduces a target-aware labeling scheme.

Dual-MRC(Mao et al., 2021) solves the ASTE task by training two MRC frameworks with shared parameters.

BART-ABSA(Yan et al., 2021) uses the BART model to extract sentiment triples in a generative manner.

RLI(Yu et al., 2023) proposes a retrieval-based ASTE approach.

4.4. Results and Analysis

The results of the main experiments are shown in Table 4. It can be observed from the results that the S&T model achieves optimal results on each data set under the F1 score.

The results of the integral experiments are shown in Table 4. It can be observed from the results that the F1 score of S&T model achieves optimal results on each dataset. Compared with the sequence labeling method, both the table-filling method and the span based method have great advantages, which proves the effectiveness of these two methods. Among them, the performance of the span based method is better, because the information in the phrase is more comprehensive, which reflects the importance of information integrity in item extraction related tasks. Both STAGE and BDTF are based on the table-filling method. It can be observed from the results of Lap14 dataset that the S&T model can cope with complex situations better than STAGE and BDTF.

4.5. Ablation Study

In order to verify the effectiveness of different modules of our S&T model, we conducted ablation studies, and the experimental results are shown in Table 5. “simple_word” means to concat the

	Sentence	Ground truth	EMC-GCN	BDTF	Ours
1	The menu is interesting and quite reasonably priced .	(menu, interesting, POS) (menu, reasonably priced, POS) (priced, reasonably, POS)	(menu, interesting, POS) (menu, ***reasonably***, POS) (priced, reasonably, POS)	(menu, interesting, POS) (menu, reasonably priced, POS) (priced, reasonably, POS)	(menu, interesting, POS) (menu, reasonably priced, POS) (priced, reasonably, POS)
2	The downstairs bar scene is very cool and chill .	(downstairs bar scene, cool, POS) (downstairs bar scene, chill, POS)	(***downstairs***, cool, POS) (***downstairs***, chill, POS) (***bar scene***, cool, POS) (***bar scene***, chill, POS)	(downstairs bar scene, cool, POS) (downstairs bar scene, chill, POS)	(downstairs bar scene, cool, POS) (downstairs bar scene, chill, POS)
3	BEST spicy tuna roll , great asian salad .	(spicy tuna roll, BEST, POS) (asian salad, great, POS)	(***tuna roll***, BEST, POS) (asian salad, great, POS)	(spicy tuna roll, BEST, POS) (asian salad, great, POS)	(spicy tuna roll, BEST, POS) (asian salad, great, POS)
4	new hamburger with special sauce is ok - at least better than big mac !	(hamburger with special sauce , ok, POS) (big mac, better than, NEG)	(***hamburger***, ok, POS) (big mac, ***better***, ***POS***)	$\emptyset$	(***hamburger with special sauce*** , ok, POS) (hamburger with special sauce , better than, NEG)
5	Service is excellent , no wait , and you get a lot for the price .	(Service, excellent, POS) (wait, no, POS)	(Service, excellent, POS) (wait, no, ***NEG***)	(Service, excellent, POS) (wait, no, POS)	(Service, excellent, POS) (wait, no, POS)
6	i love their chicken pasta cant remember the name but is sooo good.	(chicken pasta, love, POS)	(chicken pasta, love, POS) (chicken pasta, good, POS)	(chicken pasta, love, POS) (chicken pasta, good, POS)	(chicken pasta, love, POS) (chicken pasta, good, POS)

Table 6: Case studies, where erroneous elements are bolded and italicized.

span table with the simple word pair, keep the iteration module, and take the iteration output as the word embedding feature; "only_pair" indicates that only the feature representation of the span table is used to judge the label. "word_no_span" indicates that although the two feature representations are fused, the word-pair table feature representations are not converted to span representations.

Experimental results show that each module in S&T model can effectively improve the overall performance. For "simple_word", simply concat simple word-pair information cannot effectively utilize the accumulated interaction information between words. "only_pair" reflects that although the simple phrase span feature can integrate the overall information of the term, it lacks the internal association information, which affects the judgment of whether the candidate can be used as a complete term. The effect of "word_no_span" is similar to that of STAGE. Only the features of the beginning and end words is associated, which makes too much consideration for information from both ends to affect the judgment of valid segments.

4.6. Iterations and Convolution Layers

To find out the appropriate number of iterations and convolution layers of the interactive module, we conducted exploration research on Res16 and the results are shown in Figure 5.

As can be seen from the Figure 5, the model can achieve the best results when the two-layer D-GCN is iterated three times. The higher the number of iterations or convolutional layers, the worse the model performance is. It can be further concluded that for interactive module, when the number of iterations is too small, syntactic information and context information cannot be fully integrated, and too many iterations will make the model learn noise information. For the graph convolution layer, when the number of layers is too small, the learning of syntactic information is insufficient, and when the number of layers is too large, the required features may be confused.

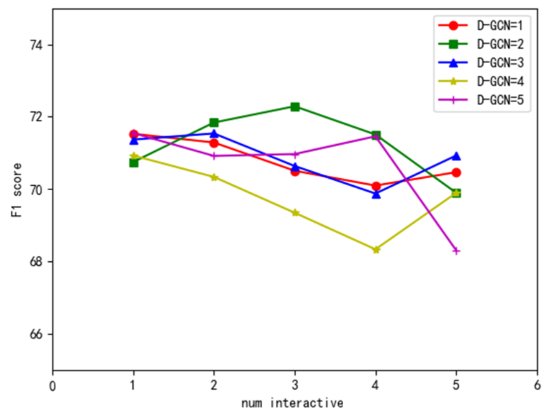

Figure 5: The exploration results of the number of iterations and the number of D-GCN layers

4.7. Case Studies

This section discusses the S&T model in this paper through the analysis results of several representative examples, as shown in Table 6. In the first example, where aspects and opinion overlap, S&T can successfully extract triplet information in this challenging task. The second and third examples show that our method produces fewer boundary errors than previous methods. In the second example, one aspect is segmented, and aspect-opinion pair to detect redundancy. In example three, there are boundary errors in the output triplets of EMC-GCN. The strategy of BDTF is to remove these error triples directly, and the S&T successfully identifies one of them. In the fifth example, although "no" is a negative word, it can be combined with "wait" to express positive polarity in the restaurant domain. The sixth example reflects the quality of the existing dataset, especially its incomplete annotation.

5. Conclusion

In this paper, we propose to combine table-filling and span to accomplish ASTE task. The table-filling method is helpful to calculate the boundary of the object item and enhance the information interaction between word pairs. Span can leverage complete terminology information to improve model performance. Span-level table-filling combines the advantages of both methods to greatly improve performance on ASTE tasks. The superiority of S&T model is demonstrated by experiments on D2 dataset.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No.62366038), and Natural Science Foundation of Inner Mongolia (No.2023MS06010). We also would like to express our sincere gratitude to the editor and two anonymous reviewers for their valuable comments, which have greatly improved this paper.

References

Hao Chen, Zepeng Zhai, Fangxiang Feng, Ruifan Li, and Xiaojie Wang. 2022a. Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction. In *Annual Meeting of the Association for Computational Linguistics*.

Shaowei Chen, Yu Wang, Jie Liu, and Yuelin Wang. 2021a. Bidirectional machine reading comprehension for aspect sentiment triplet extraction. *ArXiv*, abs/2103.07665.

Yuqi Chen, Keming Chen, Xian Sun, and Zequn Zhang. 2022b. A span-level bidirectional network for aspect sentiment triplet extraction. In *Conference on Empirical Methods in Natural Language Processing*.

Yuqi Chen, Zequn Zhang, Guangyao Zhou, Xian Sun, and Keming Chen. 2022c. Span-based dual-decoder framework for aspect sentiment triplet extraction. *Neurocomputing*, 492:211–221.

Zhexue Chen, Hong Huang, Bang Liu, Xuanhua Shi, and Hai Jin. 2021b. Semantic and syntactic enhanced aspect sentiment triplet extraction. *ArXiv*, abs/2106.03315.

Hugh A. Chipman, Edward I. George, and Robert E. McCulloch. 2008. Bart: Bayesian additive regression trees. *The Annals of Applied Statistics*, 4:266–298.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics*.

Hao Fei, Yafeng Ren, Yue Zhang, and Donghong Ji. 2021. Nonautoregressive encoder–decoder neural framework for end-to-end aspect-based sentiment triplet extraction. *IEEE Transactions on Neural Networks and Learning Systems*, 34:5544–5556.

Zhenda Hu, Zhaoxia Wang, Yinglin Wang, and Ah-Hwee Tan. 2022. Aspect sentiment triplet extraction incorporating syntactic constituency parsing tree and commonsense knowledge graph. *Cognitive Computation*, 15:337–347.

Lianzhe Huang, Peiyi Wang, Sujian Li, Tianyu Liu, Xiaodong Zhang, Zhicong Cheng, Dawei Yin, and Houfeng Wang. 2021. First target and opinion then polarity: Enhancing target-opinion correlation for aspect sentiment triplet extraction. *ArXiv*, abs/2102.08549.

You Li, Yongdong Lin, Yuming Lin, Liang Chang, and Huibing Zhang. 2022. A span-sharing joint extraction framework for harvesting aspect sentiment triplets. *Knowl. Based Syst.*, 242:108366.

Shuo Liang, Wei Wei, Xian ling Mao, Yuanyuan Fu, Rui Fang, and Dangyang Chen. 2022. Stage: Span tagging and greedy inference scheme for aspect sentiment triplet extraction. volume abs/2211.15003.

Shanshan Liu, Xin Zhang, Sheng Zhang, Hui Wang, and Weiming Zhang. 2019. Neural machine reading comprehension: Methods and trends. *ArXiv*, abs/1907.01118.

Shu Liu, Kai-Wen Li, and Zuhe Li. 2022. A robustly optimized bmrc for aspect sentiment triplet extraction. In *North American Chapter of the Association for Computational Linguistics*.

Yaxin Liu, Yan Zhou, Ziming Li, Junlin Wang, Wei Zhou, and Songlin Hu. 2023. Him: An end-to-end hierarchical interaction model for aspect sentiment triplet extraction. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31:2272–2285.

Yue Mao, Yi Shen, Chao Yu, and Longjun Cai. 2021. A joint training dual-mrc framework for aspect based sentiment analysis. *ArXiv*, abs/2101.00816.

Rajdeep Mukherjee, Nithish Kannen, Saurabh Kumar Pandey, and Pawan Goyal. 2023. Contraste: Supervised contrastive pre-training with aspect-based prompts for aspect sentiment triplet extraction. *ArXiv*, abs/2310.15577.

Haiyun Peng, Lu Xu, Lidong Bing, Fei Huang, Wei Lu, and Luo Si. 2019. Knowing what, how and why: A near complete solution for aspect-based sentiment analysis. In *AAAI Conference on Artificial Intelligence*.

Colin Raffel, Noam M. Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21:140:1–140:67.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. *ArXiv*, abs/1409.3215.

Peiyi Wang, Lianzhe Huang, Tianyu Liu, Damai Dai, Runxin Xu, Houfeng Wang, Baobao Chang, and Zhifang Sui. 2021. Explicit interaction network for aspect sentiment triplet extraction. *ArXiv*, abs/2106.11148.

Yandan Wang, Zhongtang Chen, and Shuang Chen. 2023. Es-aste: enhanced span-level framework for aspect sentiment triplet extraction. *Journal of Intelligent Information Systems*, pages 1–20.

Zhanghua Wu, Chengcan Ying, Fei Zhao, Zhifang Fan, Xinyu Dai, and Rui Xia. 2020. Grid tagging scheme for end-to-end fine-grained opinion extraction. volume abs/2010.04640.

Lu Xu, Yew Ken Chia, and Lidong Bing. 2021. Learning span-level interactions for aspect sentiment triplet extraction. In *Annual Meeting of the Association for Computational Linguistics*.

Lu Xu, Hao Li, Wei Lu, and Lidong Bing. 2020. Position-aware tagging for aspect sentiment triplet extraction. In *Conference on Empirical Methods in Natural Language Processing*.

Hang Yan, Junqi Dai, Tuo Ji, Xipeng Qiu, and Zheng Zhang. 2021. A unified generative framework for aspect-based sentiment analysis. *ArXiv*, abs/2106.04300.

Guoxin Yu, Lemao Liu, Haiyun Jiang, Shuming Shi, and Xiang Ao. 2023. Making better use of training corpus: Retrieval-based aspect sentiment triplet extraction via label interpolation. In *Annual Meeting of the Association for Computational Linguistics*.

Juntao Yu, Bernd Bohnet, and Massimo Poesio. 2020. Named entity recognition as dependency parsing. In *Annual Meeting of the Association for Computational Linguistics*.

Samson Yu, Tapas Nayak, Navonil Majumder, and Soujanya Poria. 2021. Aspect sentiment triplet extraction using reinforcement learning. *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*.

Li Yuan, Jin Wang, Liang-Chih Yu, and Xuejie Zhang. 2023. Encoding syntactic information into transformers for aspect-based sentiment triplet extraction. *IEEE Transactions on Affective Computing*.

Chen Zhang, Qiuchi Li, Dawei Song, and Benyou Wang. 2020. A multi-task learning framework for opinion triplet extraction. *ArXiv*, abs/2010.01512.

Wenxuan Zhang, Xin Li, Yang Deng, Lidong Bing, and Wai Lam. 2021. Towards generative aspect-based sentiment analysis. In *Annual Meeting of the Association for Computational Linguistics*.

Yice Zhang, Yifan Yang, Yihui Li, Bin Liang, Shiwei Chen, Yixue Dang, Min Yang, and Ruifeng Xu. 2022a. Boundary-driven table-filling for aspect sentiment triplet extraction. In *Conference on Empirical Methods in Natural Language Processing*.

Zhihao Zhang, Yuan Zuo, and Junjie Wu. 2022b. Aspect sentiment triplet extraction: A seq2seq approach with span copy enhanced dual decoder. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30:2729–2742.

Yanyan Zhao, Bing Qin, and Ting Liu. 2010. Sentiment analysis: Sentiment analysis. *Journal of Software*, 21:1834–1848.