Syntax-Based Aspect Sentiment Quad Prediction by Dual Modules Neural Network for Chinese Comments

Zhaoliang Wu[1], Shanyu Tang[2], Xiaoli Feng[1], Jiajun Zou[3] and Fulian Yin[1,4,*]

[1]College of Information and Communication Engineering, Communication University of China, Dingfuzhuang East Street No. 1, Beijing, 100024, China
[2]Cybersecurity and Criminology Centre, University of West London, St. Mary's Road, London W5 5RF, London, UK
[3]Department of Electronic Engineering, Tsinghua University, Shuangqing Road No. 30, Beijing, 100084, China
[4]The State Key Laboratory of Media Convergence and Communication, Communication University of China, Dingfuzhuang East Street No. 1, Beijing, 100024, China
*Corresponding Author: Fulian Yin. Email: yinfulian@cuc.edu.cn

Abstract: Aspect-Based Sentiment Analysis (ABSA) is one of the essential research in the field of Natural Language Processing (NLP), of which Aspect Sentiment Quad Prediction (ASQP) is a novel and complete subtask. ASQP aims to accurately recognize the sentiment quad in the target sentence, which includes the aspect term, the aspect category, the corresponding opinion term, and the sentiment polarity of opinion. Nevertheless, existing approaches lack knowledge of the sentence's syntax, so despite recent innovations in ASQP, it is poor for complex cyber comment processing. Also, most research has focused on processing English text, and ASQP for Chinese text is almost non-existent. Chinese usage is more casual than English, and individual characters contain more information. We propose a novel syntactically enhanced neural network framework inspired by syntax knowledge enhancement strategies in other NLP studies. In this framework, part of speech (POS) and dependency trees are input to the model as auxiliary information to strengthen its cognition of Chinese text structure. Besides, we design a relation extraction module, which provides a bridge for the overall extraction of the framework. A comparison of the designed experiments reveals that our proposed strategy outperforms the previous studies on the key metric F1. Further experiments demonstrate that the auxiliary information added to the framework improves the final performance in different ways.

**Keywords:** ABSA; ASQP; sentiment analysis; Chinese comments

1 Introduction

Cyber comments are widespread on the internet in China. They are rich in commercial and academic value. Through sentiment analysis and opinion mining research [1,2], users' demands can be quickly and effectively obtained. As a result, E-commerce enterprises can improve relevant products

while scholars can conduct relevant sociology or psychology research. Early sentiment analysis studies focused on entire documents or sentences [3–5]. Nevertheless, sometimes, a sentence may contain multiple sentimental expression objects and corresponding sentimental viewpoints, which may sometimes be opposite. For example, in the sentence "The food in this restaurant is good, but the service is bad," the user expresses opposite sentiments about different aspects. Traditional sentiment analysis is hard to deal with it. Therefore, Aspect-Based Sentiment Analysis (ABSA) is proposed. This task requires the identification of one or several sentimental elements. Generally, sentimental elements are the aspect term, opinion term, aspect category, and sentiment polarity. The aspect term and the opinion term are expressed in the target text. The aspect category and the sentiment polarity belong to the pre-defined category and sentiment sets, as shown in Fig. 1. In recent years, ABSA has received much attention and has been developed rapidly [6,7]. This paper focuses on the subtask Aspect of Sentiment Quad Prediction (ASQP) under ABSA of Chinese comments, which is the task for predicting the complete sentiment elements of Chinese comments.
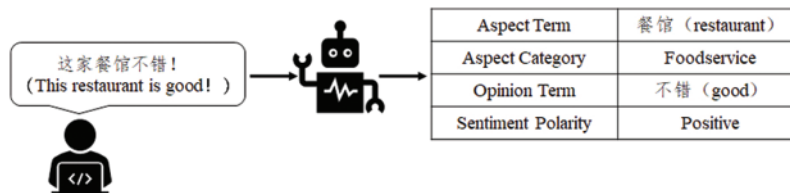
Figure 1: An example of 4 sentiment elements of ABSA

The ASQP task has attracted much attention recently due to its ability to predict the complete sentiment elements. Some researchers have noted implicit aspect terms and opinion terms in real-life comments. A concrete example is shown in Fig. 2. So, they extracted the feature representation of the whole text through the pre-training model, obtained the final result through the end-to-end method, and achieved good results [8]. Other researchers leverage the large-scale generative model to train the original comments to get a regular text containing quadruple sentiment elements information [9]. However, the objects of this research are all English texts. An important reason for this phenomenon is that labeling a sentiment quad dataset requires a large expenditure, and English is a language with abundant resources. To solve this problem, some researchers proposed transfer learning, which maps texts of different languages into a shared vector space by learning neural network models as cross-lingual learning. In recent studies, some scholars tried to use the adversarial network to map different languages and achieved good results [10]. However, these methods ignore the differences between Chinese and English. Therefore, another study has noticed that Chinese as an ideogram can be seen as a small image of Chinese characters, expanding the types of Chinese representations through multimodal learning [11]. Moreover, the Chinese phonetic alphabet can also act as auxiliary information to help neural networks to analyze emotions [12].
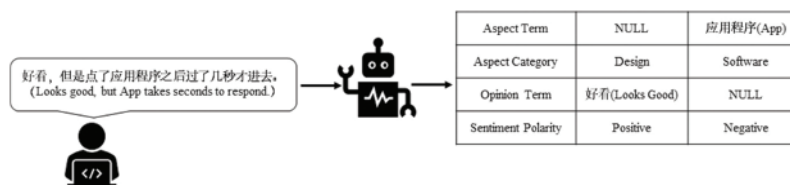
Figure 2: An example of implicit aspect and opinion in a sentence

Researchers have achieved good results in both ASQP and Chinese sentiment analysis. However, limited by the dataset, there are few pieces of research on ASQP tasks for Chinese text. In addition, Chinese ASQP research needs to pay attention to the characteristics of the Chinese language itself. With the improvement of the NLP pre-training model, the semantic information of the Chinese language has been very accurate, so we need to pay more attention to the Chinese syntax.

We propose a syntactically enhanced aspect-opinion dual extraction model. The part of speech (POS) and word dependence tree in the syntactic structure are used as auxiliary information to extract entities and entity relations, respectively. The model is divided into Aspect-Category-Sentiment Extraction (ACSE) module and the Opinion-Category-Sentiment Extraction (OCSE) module to address the above challenges. The improved performance on the baseline validates the effectiveness of our model.

The major contributions of this paper include the following:

- We utilize Chinese grammar knowledge as external knowledge to improve the model's ability to recognize entity boundaries.
- We use the Relation Attention Network to extract the relation of words in the text and add this as auxiliary information to the model for entity extraction.
- We leverage a grid tagging scheme to train the model and obtain the sentimental three-element information through the opinion and aspect extract modules. The sentimental four-element results are finally obtained through the element matching mode.

The remainder of this paper is organized as follows: Section 2 briefly gives some related complementary papers around the ABSA task. Section 3 describes our proposed model internal architecture and the specific formulas involved when it works and is trained. Section 4 states the composition structure of the dataset we used and the model's hyperparameters. Section 5 introduces a series of detailed experiments that demonstrate the effectiveness of the designed model in terms of key metrics. Furthermore, we explicate the performance of our model when dealing with cases and the types of errors through some case studies and error analysis. Finally, we conclude our primary contributions in ASQP for Chinese comments in Section 6.

2 Related Work

For ABSA tasks, the aspect term is the most central emotional element. One type of aspect term extraction is to use labels for supervised learning. The paradigm of this approach is that it maps tokens of the target text into a representational vector space and then classifies them to obtain the aspect term. The other is unsupervised aspect term extraction through rules or knowledge. In recent research, these two paradigms converge, with more accurate semantic information available through neural networks, and labeling limitations can be broken through external knowledge and rules. He et al. [13] propose an improved linear discriminant analysis model to obtain better results by embedding the text. Luo et al. [14] introduce the concept of Sememes to build neural network models using unlabeled corpus in an unsupervised framework. Yin et al. [15] and Xu et al. [16] combined the knowledge embedding with the vector representation of tokens also improved the model performance. These studies provide a way to combine external knowledge with neural network models to extract aspect terms for our research. The extraction paradigm of the aspect term still applies to the extraction of the opinion term. However, the difference is that an opinion independent of the aspect term is meaningless, so an aspect-based opinion term extraction task is proposed. The method of extracting the corresponding opinion term for a specific aspect term is generally used [17]. However, an implicit

aspect makes this method less effective, so some studies embed the whole sentence and then calculated the similarity with the specified aspect category to obtain the aspect category of implicit aspect [18–20]. These studies likewise give us insight that using aspect categories to extract the opinion term in a sentence would greatly alleviate the crippling performance effect of implicit aspects on the model.

In our research, more than entity identification is needed. We also need to match aspect terms and opinion terms. What is more, we need to focus on the categories and polarities corresponding to the entities. Some researchers have solved this problem using labels to output entities directly through the model. For example, Wu et al. [21] proposed a grid labeling scheme to build an end-to-end output model. In recent studies, scholars have tried combining grammatical, syntactic, and external affective knowledge into the model to overcome the dataset's limitations [22]. The Knowledgeable Prompt-tuning method can effectively improve the classification effect [23]. More importantly, Graph Convolutional Networks (GCN) fit well with the dependencies in syntactic structure and the structure of sentiment triads in external sentiment knowledge. Some studies have performed well by introducing GCN into entity extraction [24,25]. By learning the text structure through GCN, the model can not only associate entities before and after the temporal order but also learn the association of entities that possess dependencies. Chen et al. [26] and Feng et al. [27] used this feature of GCN in ABSA to enhance the association between sentiment elements. Inspired by this, we transformed the dependency relationships into adjacency matrices to input into our model, allowing the model to obtain information about the association between entities.

Facing the situation of the insufficient dataset in the ABSA task, besides external knowledge, it is an excellent choice to use pre-trained language models (PLMs) [28,29], which uses a large amount of unlabeled data for unsupervised learning and obtains very accurate semantic understanding. In addition, scholars also pay attention to the parameter initialization problem and use Matching features with output label-based Advanced Technique (MAT) to initialize the weights to obtain better training results [30]. Other scholars use the idea of meta-learning to make the model self-trained to achieve a one-shot effect [31]. The pre-training model in the Chinese ABSA task still provides instrumental help. Wang et al. [32] used the pre-training model to achieve SOTA results by embedding substitution for Chinese words. Earlier, Peng et al. [33] had a reasonably diverse study of ABSA in Chinese. They provided meaningful guidance on Chinese ABSA regarding syntactic structure, corpus, and other characters. They then tried to add the Chinese phonetic alphabet to the model for multidimensional feature representation and obtained good results [12]. Thus, we obtain the Chinese dependencies by PLMs to make our model better understand Chinese.

3 Methodology

3.1 Task Formulation

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a comment of n characters, let $S = \{s_{1,1}, s_{1,2}, \ldots, s_{i,j}, \ldots, s_{n,n}\}$ be the set of all possible enumerated spans in X, which i and j are the start and end of a character position for a span. Each span may be an aspect term or an opinion term of a comment. The objective of the ASQP task is to extract all possible quadruples (aspect term, opinion term, aspect category, sentiment polarity) in X, where the aspect term and the opinion term are subsets of S. The aspect category and the sentiment polarity are from pre-defined category sets C and P. There may be an overlap between aspect terms and opinion terms in this study. Scilicet, there exists the aspect category $s_{i,j}$ and the opinion term $s_{k,l}$ where $i < k < j$.

3.2 Model Architecture

The pre-trained model can effectively extract a piece of text's semantic and grammatical information, so we leverage the BERT to embed the text to obtain the characters' representation. At the same time, the Stanford CoreNLP tool obtains the POS of words and the dependency tree of sentences in the text. The next process is as follows: first, the dependency tree is transformed into an adjacency matrix, which is then put into the Relation Extraction Module (REM) along with the character representation to obtain relational information in the sentence. Second, the relation information, with character representation and POS, is input to the ACSE module and OCSE module, respectively, to obtain Aspect-Category-Sentiment triples and Opinion-Category-Sentiment triples. Finally, the model matches the two sentiment elements of aspect category and sentiment polarity in Aspect-Category-Sentiment triples and Opinion-Category-Sentiment triples and obtains the Aspect Sentiment Quad by multivariate inference. The Overview of our model is shown in Fig. 3.
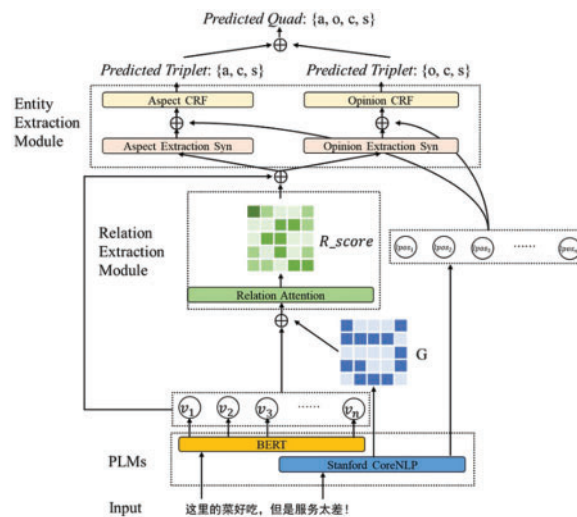

Figure 3: Overview of our proposed model

3.2.1 Character Representation from BERT

We use the BERT language model [28] as the embedding method for tokens:

$$\{v_1, \ldots, v_n\} = BERT \left(\{c_1, \ldots, c_n\}\right) \tag{1}$$

where v_i is the output representation for the character c_i.

3.2.2 POS and Dependency Tree from Stanford CoreNLP

Stanford CoreNLP[1] is a set of NLP tools developed by the Stanford NLP Group based on their scientific work. It provides various tools to process natural language and can achieve stem reduction and tag the POS of words. For the ABSA task, POSs contribute different aspect terms and opinion terms. The aspect term is composed of nouns, while the opinion term is composed of adjectives. The dependency tree also has an auxiliary role for the ABSA task. The opinion term is dependent on the aspect term. Therefore, we leverage Stanford CoreNLP as a pipeline to generate POS and dependency trees of words:

---

[1] https://stanfordnlp.github.io/CoreNLP/

$$\{pos_1, \ldots, pos_m\} = \text{CoreNLP}\,(\{w_1, \ldots, w_m\}) \tag{2}$$

$$G = \text{CoreNLP}\,(\{w_1, \ldots, w_m\}) \tag{3}$$

where pos_i is the output POS for the word w_i, G is the output adjacency matrix of the sentence.

There are some differences between Chinese and English in terms of semantic units. In Chinese, the smallest ideographic unit is a word, a combination of one or more characters [34]. For example, the Chinese word "参差不齐" is combined with four characters: "参," "差," "不," and "齐." CoreNLP adopts the method of first segmenting and then classifying Chinese text, so the smallest unit of Chinese text it processes is a word. In contrast, BERT is embedded for each character because the same character has different semantics in different word combinations. Then a problem arises. The same sentence will have different length outputs in CoreNLP and BERT. Taking the above example, the phrase "参差不齐" has four outputs corresponding to four characters in BERT, while only one output corresponds to one word in CoreNLP. To match the output results, we marked the output of CoreNLP with "BIES," added "BIES" to the attributes of the generated words, and split them into word outputs. Still taking the phrase "参差不齐" as an example, the initial output of this phrase through CoreNLP should be one output: "VV" after being marked by "BIES," the output is four: B-VV, I-VV, I-VV, E-VV.

$$\{lpos_1, \ldots, lpos_n\} = \text{Label}\,(\{pos_1, \ldots, pos_m\}) \tag{4}$$

3.2.3 Relation Extraction Module

The Aspect Sentiment Quad task not only needs to identify aspect terms and opinion terms but also needs to match them. For the comment: "披萨好吃但是服务很差(This pizza is delicious, but the service is terrible)." There are two aspect terms {披萨(pizza); 服务(service)} and two opinion terms{好吃(delicious); 很差(terrible)}, there is a dependency relationship between them, where "delicious" is the dependency to "pizza" while "terrible" is the dependency to "service." According to this feature, we set up the Relation Extraction Module to extract the relationship in the text as the "word pair." They are input into subsequent modules.

In this module, the text's adjacency matrix G is fed into the Relation Attention Network along with the semantic representation V. The output of this module is the correlation matrix R_score.

$$G = \left\{ g_{1,1}, \ldots g_{i,j}, \ldots, g_{n,n} \right\} \tag{5}$$

$$R\_h_i = \frac{\sum_{j=1}^{n} v_i g_{i,j}}{\sum_{j=1}^{n} g_{i,j}} \tag{6}$$

$$R\_score = Attention\,(tanh\,(R\_H \cdot W_{r1} + V \cdot W_{r2})) \tag{7}$$

where R_h_i is the intermediates output in this module, and R_H is the matrix of R_h_i.

3.2.4 Entity Extraction Module

Entity Extraction module contains ACSE module and OCSE module. They have the same neural network structure, and during training, we give these two modules different inputs and labels, thus making the two modules have different parameters. Since the opinion term has a dependency on the aspect term, the correlation matrix R_score obtained by the Relation Extraction Module improves the extraction accuracy of the module by weighting the inputs:

$$A\_r_i = softmax\left(\sum_{j=1}^{n} r_{i,j}\right) \tag{8}$$

$$O\_r_j = softmax \left(\sum_{i=1}^{n} r_{i,j} \right) \tag{9}$$

$$A\_H = tanh \left(A\_R \cdot V \cdot W_{A1} + V \cdot W_{A2} \right) \tag{10}$$

$$O_H = tanh \left(O_R \cdot V \cdot W_{O1} + V \cdot W_{O2} \right) \tag{11}$$

where $r_{i,j}$ is the dependency score of the i-th position on the j-th position in y R_score. A_R and O_R are matrices of A_r_i and O_r_j, respectively.

Then, we combine the POS information with A_H and O_H through the attention mechanism [35] and finally get the final label corresponding to each word through Conditional Random Field (CRF) [36]:

$$A\_label = CRF \left(Attention \left(A\_H, lpos \right) \right) \tag{12}$$

$$O\_label = CRF \left(Attention \left(O\_H, lpos \right) \right) \tag{13}$$

The *A_label* contains {*aspect term, categories, sentiment polarities*}, while *O_label* contains {*opinion term, categories, sentiment polarities*}. Finally, we perform multivariate inference on and to obtain the sentiment quad. Specifically, the inference information {categories, sentiment polarities} in the two labels is the primary basis, and the related information in the relationship extraction module is the secondary information. Then, according to the task's requirements, we perform two types of processing. In the task which does not consider implicit aspects and opinion, we only output the matched labels. While in the task that considers implicit aspect and opinion, we treat the unmatched labels as separate quads, and the corresponding entity elements are marked as NULL.

3.3 Training

During training, given an input sentence s with manually annotated label quad:

$$l = \{\widehat{l_k} \left(\hat{a}, \ \hat{o}, \ \hat{c}, \hat{s} \right) \}_{k=1}^{K} \tag{14}$$

where $\hat{a}$, $\hat{o}$, $\hat{c}$, and $\hat{s}$ are the manually annotated aspect term, opinion term, categories, and sentiment polarities. K means that there are K quad ancestors for s. Due to the existence of implicit aspect and implicit opinion, $\hat{a}$ and $\hat{o}$ may be null.

We calculate the joint loss value by calculating the output of the three modules with the gold label:

$$loss = \omega_1 \mathcal{L}_r \left(R\_score, \ L^{(r)} \right) + \omega_2 \mathcal{L}_e \left(A\_label, \ L^{(a)} \right) + \omega_3 \mathcal{L}_e \left(O\_label, \ L^{(o)} \right) \tag{15}$$

where ω is the weight of the loss function. $L^{(r)}$ is the gold adjacency matrix of aspect term and opinion term, $L^{(a)}$ and $L^{(o)}$ are gold ACS triples and OCS triples. $\mathcal{L}_r$ is the cross-entropy loss function, while $\mathcal{L}_e$ is the negative log-likelihood loss.

$$\mathcal{L}_r = \frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} - r_{i,j} log l_{i,j}^{(r)} - \left(1 - r_{i,j} \right) \log \left(1 - l_{i,j}^{(r)} \right) \tag{16}$$

$$\mathcal{L}_e = log \left(\sum_{\hat{y} \in label} e^{s(x,y)} \right) - s \left(x, y \right) \tag{17}$$

$$s \left(x, y \right) = \sum_{i=0}^{l} T \left(y_i + y_{i+1} \right) + \sum_{i=0}^{l} U \left(x_i + y_i \right), \tag{18}$$

where $\hat{y}$ is the gold label, x is the input of CRF, y is the output of CRF, l is sequence length, T and U are parameters, $T\,(y_i + y_{i+1})$ is transfer scores where the i step label is y_i, while $(i + 1)$ step label is y_{i+1}, $U\,(x_i + y_i)$ is the launch scores of the i step input being x_i.

4 Datasets and Experiment Setups

4.1 Datasets

We collected the comments on JD e-commerce under agricultural products, created a labeling platform, and recruited volunteers to label each comment with a sentiment quad. The volunteers set the number of comments' quads according to their understanding. In each quad, the aspect and opinion terms are selected from the comment, and the option can be left blank if there is an implicit case. The aspect category and sentiment polarity are selected from the items already set. When a volunteer finished labeling a message, another volunteer checked this label, which would be treated as a gold label after confirming it was correct. A total of 8670 comments were labeled, and we randomly divided them into 7000 and 1670 as train and test sets. We count explicit aspect (EA), explicit opinion (EO), implicit aspect (IA), and implicit opinion (IO), respectively. The details are shown in Table 1.

Table 1: Statistics of our quad datasets

	Train	Test
Sentences	7000	1670
EA & EO	9555	2307
IA & EO	5093	1235
EA & IO	64	19
All	14712	3561

4.2 Evaluation Metrics and Implementation Details

We used F1-score to evaluate the performance of the model. We consider a predicted Aspect Sentiment Quad is correct if the gold standard annotations are completely consistent with the prediction.

BERT representation V has 768 dimensions, while POS embedding is set to 100 dimensions, and the hidden output in three modules is 250 dimensions. We adopted the Adam optimizer with an initial learning rate of $1e - 3$, with subsequent 10 percent decay every ten epochs. We set dropout to 0.5 and 10 batch size while training and 50 batch size while testing. When F1-score does not improve for three consecutive epochs, it triggers an early stop to stop training. We leveraged five seed numbers for training, including the main experiment, the comparison experiment, and the ablation experiment, and the obtained results were averaged as the final results.

5 Results and Discussion

5.1 Performance Evaluation

Our baselines are divided into pipelines [8] and unified methods [9]. Pipeline methods use Extract-Classify-ACOS to co-extract the aspect-opinion and then predict the category-sentiment given the extracted aspect-opinion pairs. Unified methods perform joint extraction of all aspect-level sentiment elements: aspect terms, aspect categories, opinion terms, and sentiment polarities. We mainly modify

the PARAPHRASE modeling paradigm to cast the ASQP task and the different situations where implicit aspects and implicit opinions are considered. Since they can only simultaneously extract the quad containing implicit aspect and implicit opinion, we modify the two models accordingly. We also modify our model to ensure that the model output is consistent with the baseline. Then we got the results for the ASQP task. P, R, and F1 denote precision, recall, and F1-score, respectively. IA and IO denote implicit aspect and implicit opinion, respectively. Scores are averaged over five runs with different seeds, as shown in Table 2.

Table 2: Experimental results for ASQP

Model	With four elements			Consider IA & IO		
	P	R	F1	P	R	F1
ASQP as paraphrase generation	37.18	36.55	36.85	23	22.62	22.81
Extract-Classify-ACOS	23.82	**41.25**	30.11	18.53	34.01	23.95
Ours	**42.66**	39.21	**40.81**	**33.49**	**34.52**	**33.94**

We obtain the following important observations from the comparison results shown in Table 2. First, our model outperforms all other models on the primary metric, F1. Second, our model is better than other models in precision (more than 5% improvement). However, in a recall, our model is lower than other models in the task without considering implicit aspects and opinions. This result means that our model is doing more output than other models. Third, our model achieves a more substantial performance when implicit aspect and implicit opinion are considered than when they are not. Although our model also considers the association between entities in extraction, the association between entities is not a decisive factor, so there is a more extensive improvement when implicit aspects and implicit opinions are considered here.

Considering that there are currently a few models for sentiment quadruple extraction, we also compared the model's performance with other ABSA tasks. We obtained the main results for the ATE task and E2E-ABSA task. Scores are averaged over five runs with different seeds. The results are listed in Table 3.

Table 3: Experimental results for other ABSA tasks

Task	ATE			E2E-ABSA		
Model	P	R	F1	P	R	F1
BERT + Linear	72.63	76.91	74.70	70.85	70.85	70.79
BERT + GRU	73.09	78.69	75.77	70.47	72.12	71.27
BERT + SAN	72.16	78.24	75.06	71.07	71.64	71.32
BERT + TFM	71.31	79.28	75.06	69.96	72.33	71.11
BERT + CRF	72.79	77.11	74.87	**71.15**	70.81	70.96
Ours	**77.01**	**83.98**	**79.38**	69.93	**75.59**	**71.77**

In order to maintain the consistency of model training, whether in ATE or E2E-ABSA tasks, the label of our model is still a joint label, and the label contains {aspect, category, sentiment} three-element information. However, the baseline model does not use the same label because we meet the requirements of their original label. In the ATE task, they are single-label, and in the E2E-ABSA task, they are {aspect, sentiment} joint-label. In theory, the performance of our model would improve. Our model uses the same label as the baseline.

Compared with other models, our model improves the key indicator F1 in both the ATE task and the E2E-ABSA task, and the improvement on the ATE task is more significant than that of the E2E-ABSA task. This result means that our model improves more in entity extraction.

5.2 Ablation Study

In order to verify the effectiveness of different vital components of our model, we designed three variants to study the impacts of various architectural decisions on the model's performance and received the main results. Scores are averaged over five runs with different seeds. The results from ablations studies are shown in Table 4.

Table 4: Results of ablation experiment for ASQP

Model	With four elements			Consider IA & IO		
	P	R	F1	P	R	F1
w/o POS	42.29	37.62	39.77	32.88	33.26	33.05
w/o dependency trees	42.51	38.79	40.43	**33.67**	33.89	33.71
w/o relation	39.88	39.95	39.91	32.08	34.71	33.34
ASQP for Chinese comment	**42.66**	**39.21**	**40.81**	33.49	**34.52**	**33.94**

POS and dependency trees are input into the model as auxiliary information for extracting the four sentiment elements, while relation is the output of the Relation Extraction Module, and it is also input into the Entity Extraction module as auxiliary information. It can be seen from the results that when these are removed, the performance of the model decreases. As tokens of different POSs contribute differently to the aspect entity and the opinion entity, it can help the model identify aspect and opinion terms. In addition, the PLMs have already divided the original sentence into tokens and fed the model some auxiliary information during embedding. The dependency trees and Relation Extraction Module provide the model with the possibility of relationships between entities, which can help the model combine the aspect and opinion terms. When the model does not consider the implicit aspect and the implicit opinion, that is, when there is a relationship between entities in the sentence, dependency trees effectively provide the relationship between entities and improve the model's performance. Therefore, removing dependency trees has high precision after considering the implicit aspects and opinions.

In order to further explore the influence of different components on the final result, we designed four other experiments, namely Aspect-Category-Sentiment Detection (ACSD), Opinion-Category-Sentiment Detection (OCSD), Opinion Term Extraction (OTE), and Aspect Term Extraction (ATE). Their results are shown in Fig. 4. ATE and OTE are single ABSA tasks and entity extraction tasks. In both tasks, the model with all its components still performs best. However, in the two tasks of OTE and ATE, different ablation experiments showed different performance degradation. In the ATE task, the goal is to capture Aspect terms. It is strongly associated with POS. For example, many of the Aspect

Terms are nouns. So, when we input POS information into the model, grasping the aspect term would be improved. However, in the OTE task, the opinion term is often a clause rather than a single word in the Chinese text, so the improvement of POS for this task is limited.
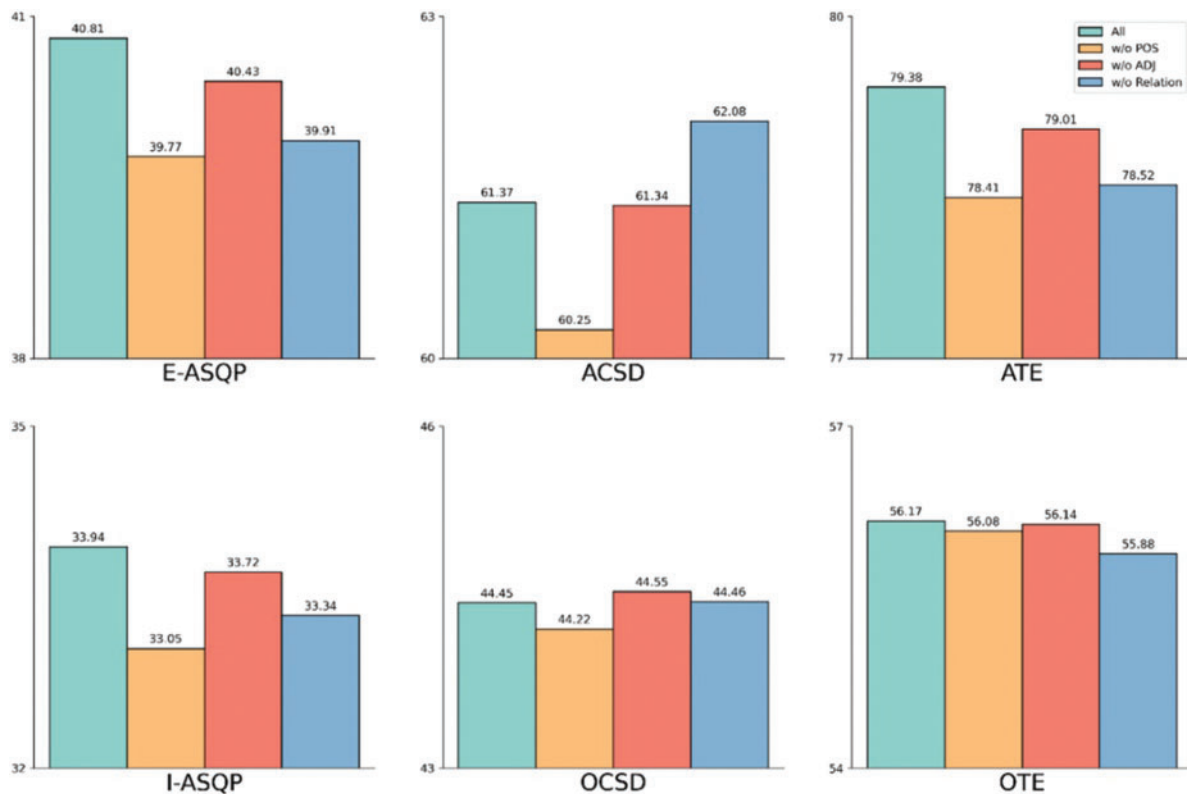

Figure 4: Multi-task ablation experiment

The relation extraction module of the model, like the adjacency matrix (ADJ) transformed from the dependency tree, provides the dependency relationship between words in sentences for the model. However, the difference is that the adjacency matrix provides the prior knowledge, while the dataset trains the relation extraction module. From the results, prior knowledge can provide more information to improve the model's performance in most cases. Simultaneously, we also notice that the model's performance without the relation extraction module in the ACSD task is better, while without word relation information in the OCSD task performs well. Compared with the single ABSA task, the ACSD task and the OCSD task need to infer the Aspect Category and Sentiment Polarity based on entity extraction. In our model, we aggregate the vector representations of related tokens and then classify them. This step increases the redundant information of the tokens to be predicted and interferes with the model's judgment. Nevertheless, the cost is worth it. Because although this step reduces the model's ability to reason about entities, it can be seen in the final ASQP task that this step increases the degree of association between the two entities (aspect terms and viewpoint terms).

5.3 Case Study

To better analyze under what circumstances our model would fail, we put the test set into the model for prediction and compared the final results with the gold label. We count the different types

of mistakes in the entity prediction, as shown in Fig. 5. It can be seen that the extracting ability of the model for the aspect term is significantly higher than that of the opinion term.
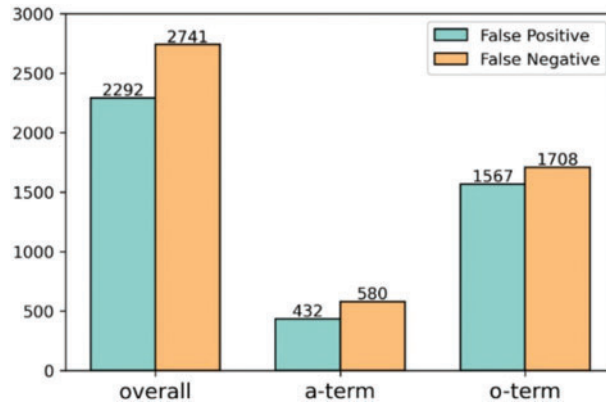
Figure 5: Number of quads mistakes due to entity extraction mistakes

In order to find the reasons, we extracted some prediction results and carried out a case study with the gold label. Part of the content is shown in Table 5. We found some incorrect results due to language features. The first is the accuracy of the expressions of some words. For example, in Example 1 in Table 5, the token "性价比 (cost-effective)" means both price and quality, so "性价比不高 (Cost-effective is not high)" not only means that the price is too high but also means that the quality is too low. Our model only considers one situation attributed to the aspect category of price. Secondly, there are differences in Chinese token segmentation in the Opinion term. For example, in example 2 in Table 5, the entire sentence can be used as the Opinion term. Unlike English, although these viewpoints can be segmented, they are still expressions of a viewpoint as a whole. However, this leads to the overlap of viewpoint and aspect words, making it difficult for the model to identify them because the model prefers to segment the whole sentence into smaller parts to identify them without a global view. Thirdly, because grammar usage in online Chinese comments is casual, extracting emotional elements using the rule-constrained model is more challenging. For example, in example 2 in Table 5, the model incorrectly sets the opinion term on all possible aspect terms.

Table 5: Comparison table of prediction results

Example-1	
Sentence:	'有点少...性价比不高!口感一般'
	(a bit small. . . Cost-effective is not high! General taste)
Gold label:	**('Cost-effective', 'not high', 'other', 'NEG'),**
	('taste', 'General', 'mouthfeel', 'NEU'),
	(−1, 'a bit small', 'wight', 'NEG').
Prediction:	**('Cost-effective', 'not high', 'price', 'NEG')×,**
	('taste', 'General', 'mouthfeel', 'NEU')√,
	(−1, 'a bit small', 'wight', 'NEG')√.
Example-2	
Sentence:	'好吃,味道不错,口感饱满,'

(Continued)

Table 5: Continued

Example-1	
Gold label:	(Delicious, good taste, full mouthfeel,) ('taste', 'Delicious, good taste, full mouthfeel', 'mouthfeel', 'POS')
Prediction:	**('taste', 'Delicious', 'mouthfeel', 'POS') ×,** **('mouthfeel', 'good', 'mouthfeel', 'POS') ×,** **('taste', 'good', 'mouthfeel', 'POS') ×,** **('mouthfeel', 'good', 'mouthfeel', 'POS') ×,** **('taste', 'full', 'mouthfeel', 'POS') ×,** **('mouthfeel', 'full', 'mouthfeel', 'POS') ×,**

Finally, to explore the model's reasoning ability, we counted the number of errors in the prediction results caused by aspect categories and the sentiment polarity, as shown in Fig. 6.

Overall, the reasoning ability of the model is stronger than the entity extraction ability, and the judgment ability of the emotion polarity is stronger than the aspect classification ability. From the previous case study, it can also be seen that the sentiment polarity judgment of Chinese text is relatively straightforward, but there are ambiguities in the classification of aspects.
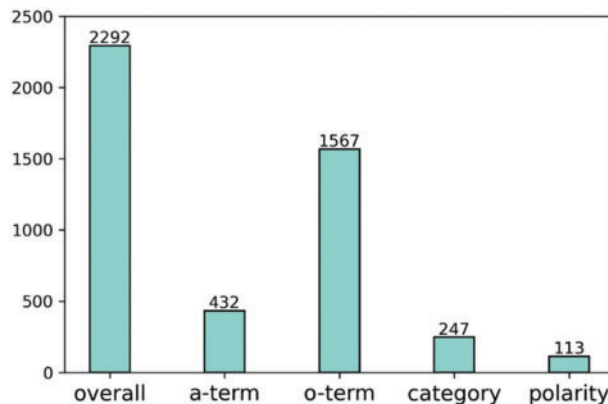
Figure 6: Number of quads w.r.t the mistake type

5.4 Limitation

As with most studies, there are limitations to the design of our current study. On the one hand, the model's performance is limited because the ASQP task needs to extract many elements, but the size of our dataset is limited. On the other hand, the delineation results are subjective because there is no clear definition of entity boundaries. Therefore, especially in the Chinese version of the opinion term, it is straightforward to cause overfitting, which leads to the degradation of model performance. Furthermore, the single prediction evaluation metric must not perfectly reflect the subjective delineation results.

6 Conclusion and Future Work

A comprehensive and accurate exploration of cyber users' opinions about products is critical to finding users' interests. This paper proposed a fusing Chinese POS and sentence structure-based neural network model for sentiment quadruple extraction from Chinese comments. Contrasting with previous methods, we integrated Chinese POS and syntactic structure into the neural network, which improved the ability of entity extraction and the extraction relationship between entities. In addition, we combined abstract sentiment elements with entities through labels so that the model achieves the ability of sentiment quad extraction. Experiments were conducted on a real-world dataset, and a significant improvement in metrics performances validates the effectiveness of our model. For future work, we will target our limitations. Firstly, we will introduce a sentiment knowledge graph into the study to compensate for the size of the dataset. Secondly, we will consider using diversified entity prediction criteria. For example, we can utilize similarity calculations to judge the gap between the extracted entities and the gold label.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[2] M. A. Adnan and A. Mosavi, "Aspect level songs rating based upon reviews in English," *Computers, Materials and Continua*, vol. 74, no. 2, pp. 2589–2605, 2022.

[3] P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *ACL*, Philadelphia, PA, USA, pp. 417–424, 2002.

[4] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques," in *EMNLP*, Philadelphia, PA, USA, pp. 79–86, 2002.

[5] H. Yu and V. Hatzivassiloglou, "Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences," in *EMNLP*, Sapporo, Japan, pp. 129–136, 2003.

[6] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, 2016.

[7] A. Nazir, Y. Rao, L. Wu and L. Sun, "Issues and challenges of aspect-based sentiment analysis: A comprehensive survey," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 845–863, 2020.

[8] H. Cai, R. Xia and J. Yu, "Aspect-category-opinion sentiment quadruple extraction with implicit aspects and opinions," in *ACL-IJCNLP*, Bangkok, THA, pp. 340–350, 2021.

[9] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing *et al.*, "Aspect sentiment quad prediction as paraphrase generation," in *EMNLP*, Punta Cana, DOM, pp. 9209–9219, 2021.

[10] W. Wang and S. J. Pan, "Transition-based adversarial network for cross-lingual aspect extraction," in *IJCAI*, Stockholm, SWE, pp. 4475–4481, 2018.

[11] Y. Meng, W. Wu, F. Wang, X. Li, P. Nie *et al.*, "Glyce: Glyph-vectors for Chinese character representations," in *NeurIPS*, Vancouver, CAN, pp. 32, 2019.

[12] H. Peng, Y. Ma, S. Poria, Y. Li and E. Cambria, "Phonetic-enriched text representation for Chinese sentiment analysis with reinforcement learning," *Information Fusion*, vol. 70, pp. 88–99, 2021.

[13] R. He, W. S. Lee, H. T. Ng and D. Dahlmeier, "An unsupervised neural attention model for aspect extraction," in *ACL*, Vancouver, CAN, pp. 388–397, 2017.

[14] L. Luo, X. Ao, Y. Song, J. Li, X. Yang *et al.,* "Unsupervised neural aspect extraction with sememes," in *IJCAI*, Macao, China, pp. 5123–5129, 2019.

[15] Y. Yin, F. Wei, L. Dong, K. Xu, M. Zhang *et al.,* "Unsupervised word and dependency path embeddings for aspect term extraction," in *IJCAI*, New York, NY, USA, pp. 2979–2985, 2016.

[16] H. Xu, B. Liu, L. Shu and P. S. Yu, "Double embeddings and CNN-based sequence labeling for aspect extraction," in *ACL*, Melbourne, AUS, pp. 592–598, 2018.

[17] Z. Wu, F. Zhao, X. Dai, S. Huang and J. Chen, "Latent opinions transfer network for target-oriented opinion words extraction," in *AAAI*, New York, NY, USA, pp. 9298–9305, 2020.

[18] T. Shi, L. Li, P. Wang and C. K. Reddy, "A simple and effective self-supervised contrastive learning framework for aspect detection," in *AAAI*, Virtual Conference, pp. 13815–13824, 2021.

[19] S. Tulkens and A. van Cranenburgh, "Embarrassingly simple unsupervised aspect extraction," in *ACL*, Seattle, WA, USA, pp. 3182–3187, 2020.

[20] A. Banjar, Z. Ahmed, A. Daud, R. A. Abbasi and H. Dawood, "Aspect-based sentiment analysis for polarity estimation of customer reviews on twitter," *Computers, Materials & Continua*, vol. 67, no. 2, pp. 2203–2225, 2021.

[21] Z. Wu, C. Ying, F. Zhao, Z. Fan, X. Dai *et al.,* "Grid tagging scheme for aspect-oriented fine-grained opinion extraction," in *Findings of EMNLP*, Punta Cana, DOM, pp. 2576–2585, 2020.

[22] S. Wu, H. Fei, Y. Ren, D. Ji and J. Li, "Learn from syntax: Improving pair-wise aspect and opinion terms extraction with rich syntactic knowledge," in *IJCAI*, Virtual Conference, pp. 3957–3963, 2021.

[23] S. Hu, N. Ding, H. Wang, Z. Liu, J. Li *et al.,* "Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification," in *ACL*, Dublin, IRL, pp. 2778–2788, 2022.

[24] A. P. B. Veyseh, N. Nouri, F. Dernoncourt, D. Dou and T. H. Nguyen, "Introducing syntactic structures into target opinion word extraction with deep learning," in *EMNLP*, Punta Cana, DOM, pp. 8947–8956, 2020.

[25] S. Mensah, K. Sun and N. Aletras, "An empirical study on leveraging position embeddings for target-oriented opinion words extraction," in *EMNLP*, Punta Cana, DOM, pp. 9174–9179, 2021.

[26] H. Chen, Z. Zhai, F. Feng, R. Li and X. Wang, "Enhanced multi-channel graph convolutional network for aspect sentiment triplet extraction," in *ACL*, Dublin, IRL, pp. 2974–2985, 2022.

[27] A. Feng, Z. Gao, X. Song, K. Ke, T. Xu *et al.,* "Modeling multi-targets sentiment classification via graph convolutional networks and auxiliary relation," *Computers, Materials & Continua*, vol. 64, no. 2, pp. 909–923, 2020.

[28] J. Devlin, M. Chang, K. Lee and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *NAACL-HLT*, Las Vegas, NEV, USA, pp. 4171–4186, 2019.

[29] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi *et al., Roberta: A Robustly Optimized BERT Pretraining Approach*. Los Alamos, NM, USA: arXiv, 2019. [Online]. Available: https://arxiv.org/pdf/1907.11692.pdf

[30] A. Thakkar, D. Mungra, A. Agrawal and K. Chaudhari, "Improving the performance of sentiment analysis using enhanced preprocessing technique and artificial neural network," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1771–1782, 2022.

[31] K. He, R. Mao, T. Gong, C. Li and E. Cambria, *Meta-Based Self-Training and Re-Weighting for Aspect-Based Sentiment Analysis*, Manhattan, NY, USA: IEEE Transactions on Affective Computing, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9870538

[32] J. Wang and N. Li, "BiLSTM-ATT Chinese sentiment classification model based on pre-training word vectors," in *MLKE*, Guilin, China, pp. 245–250, 2022.

[33] H. Peng, E. Cambria and A. Hussain, "A review of sentiment analysis research in Chinese language," *Cognitive Computation*, vol. 9, no. 4, pp. 423–435, 2017.

[34] J. L. Packard, "Morphology: Morphemes in Chinese," in *The Oxford Handbook of Chinese Linguistics*, Oxford, UK: Oxford Handbooks, pp. 262–273, 2015.

[35] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. N. Jones *et al., Attention is All you Need*. Los Alamos, NM, USA: arXiv, 2017. [Online]. Available: https://arxiv.org/pdf/1706.03762.pdf

[36] J. D. Lafferty, A. K. Mccallum and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *ICML*, Williamstown, MA, USA, pp. 282–289, 2001.