A Transformer-based Approach for Identifying Target-oriented Opinions from Travel Reviews

Haoda Qian[†1,2] Zaichuan Tang[†1,2] Yajun Ren[1,2] Qiudan Li*[1] Daniel Zeng[1,2]
[1]The State Key Laboratory of Management and Control for Complex Systems
Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China
{qianhaoda2019, tangzaichuan2019, renyajun2021, qiudan.li, dajun.zeng}@ia.ac.cn

***Abstract*—Performing target-oriented opinion word extraction (TOWE) from online travel reviews is a valuable reference for both tourists and attraction administration department. This paper formulates a novel research topic of identifying target-opinion pair from Chinese travel review corpus. Learning target-oriented representation accurately, locating the opinion word and extracting the complete opinion are three major challenges. Hence, we leverage aspect-based query, pos-tag and relative position and devise appropriate structure to fuse them in an encoder-decoder framework. Specifically, in the encoder, the target-fused (aspect, review) pair and the pos-tag label are encoded by transformers to model the global dependency, in the decoder, a BiLSTM is adopted to enhance contextual representation by incorporating relative position information. A real-world Chinese travel dataset for TOWE task is constructed, and the experimental results demonstrate the efficacy of the proposed model. Extensive ablation experiments are also conducted to study the effect of different components of the model.**

***Keywords*—Chinese Tourist Review; Opinion Mining; Transformers**

I. INTRODUCTION

As travel has become a major way to enrich life, online travel review sites are increasingly becoming important channels for people to share travel experience and express their opinions on attractions. Users with travel plans usually spend a lot of time and effort seeking advice by reading online travel reviews. Therefore, mining fine-grained opinions from massive tourism reviews is of great practical value for both tourists and scenic spot management departments. Fig. 1 shows an example of a travel review, presented by the original simplified Chinese and its corresponding English translation. "The view from the top of the mountain is too beautiful to be absorbed at all once, the zip wire is very fun, but the traffic is a bit inconvenient" is the review text, "too beautiful to be absorbed at all once (美不胜收)" is the opinion word that describes "the view from the top of the mountain (山顶景色)", "very fun (很好玩)" is the opinion word for "zip wire (滑索)", " a bit inconvenient (有点点不便利)" is the opinion word for "the traffic (交通)". By identifying target-oriented opinions, users can quickly know about the attraction and traffic information, thus make a wiser travel plan, management department may organize regular shuttle services to improve the traffic environment.

Previous research about target-oriented opinion words extraction (TOWE) mainly focuses on English laptop and restaurant reviews. Several challenges need to be addressed for travel scenario. First, the expression of opinion words in travel reviews is more diverse, there exist more complex grammatical rules, even include Chinese idioms, it is difficult to identify complete opinion phrases with degree descriptions. In the above mentioned example, "too beautiful to be absorbed at all once (美不胜收)" is a Chinese idiom that expresses a very beautiful feeling, and " a bit inconvenient (有点点不便利)" is a slightly dissatisfied complaint about the traffic conditions, the strength of opinions may help users gain a deeper understanding of the advantage and disadvantage of the attractions, thus make better consumption choices. Second, given different targets, the model needs to provide different opinion words for the same review. Extracting such opinion words requires the model to integrate useful Chinese linguistic knowledge, and learn target-specified contextual representation.

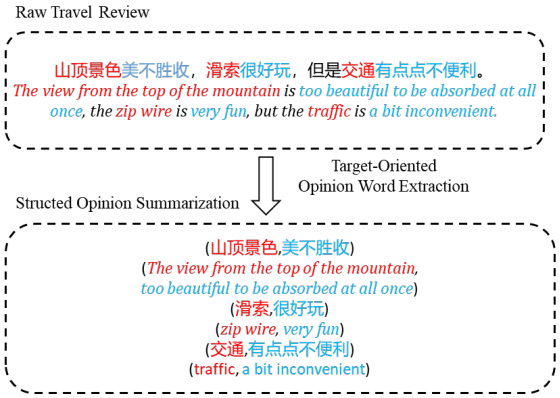

Fig. 1 An example of travel review sentence

Existing studies have shown that linguistic knowledge [1, 2] and relative position information [3, 4] play a very important role in information extraction. However, little research has systematically investigated these two kinds of information to perform target concerned opinion words mining on Chinese travel reviews. In the process of contextual representation learning, the pos tags of words can help divide the boundaries of Chinese words, and reveal the dependencies between words, thus capturing degree adverbs and transition conjunctions to form a more complete opinion. In addition, the relative position information can help locate the opinion. In this paper, we propose a *Transformer-based Multi-level Model* that combines these information with pretrained BERT to learn a task-aware representation from limited training samples. Specifically, the model uses the BERT weight as the initialization and adopts a hierarchical manner to integrate

[†] Equal Contribution
*Corresponding Author

domain and task knowledge. In the encoding layer, the task-aware (aspect, review) pair and the pos-tag label are used as the input of BERT, in the decoding layer, a BiLSTM integrates the relative position with contextual representation to model the sequential relationship between words.

We empirically evaluate the performance of the proposed model on a newly constructed real-world Chinese travel review dataset. Experimental results show that taking into account the pos-tag and relative position information could allow for more accurate target-oriented opinion words extraction. We also conduct extensive ablation analysis on modules of the model. The observed patterns could help gain deep insights into the public feedback of the attraction, thus make better decisions.

In summary, our main contributions are as follows:

- We propose a transformer-based model that fuses target-specified information, pos-tag information as well as relative position in a multi-level manner to capture task-aware patterns.
- We demonstrate the effectiveness of the model on a real-world Chinese travel review dataset using quantitative and qualitative experiments.

The rest of this paper is organized as follows: In section II, we review related works. In section III, we present the details of our proposed model. The experiments are discussed in section IV. Finally, we summarize our work and put forward future research directions in section V.

II. Literature Review

Our work is related to target-oriented opinion words extraction, sentiment analysis in travel domain. In this section, we review the related works.

A. Target-oriented opinion words extraction

Many research works have focused on aspect term and opinion term extraction separately [5-8]. However, these works rarely consider the semantic correlation between aspect and opinion. Such correlation leads to the birth of target-oriented opinion extraction (TOWE) task. This task aims to obtain opinion phrases that describe a specified target by learning the representation of the target. [9] first proposed the task and developed Inward-Outward LSTM model that passed the target information to the left context and right context, respectively, and then combined these context information and global context to encode the representation of sentence, [10] adopted transfer learning to solve the challenge of low resource, which transfers the latent opinion information from the pre-trained sentiment analysis model to the TOWE model, [11] constructed a syntactic dependency tree and used GCN to mine the dependencies among words, [12] empirically found that BiLSTM is useful for modeling relative position information and GCN can identify word dependencies, [13] added [SEP] identifiers on both sides of the target, and obtained the target-based representation through multi-layer Transformer encoding (BERT), and then extracted the opinions by means of sequence annotation.

The above work mainly focuses on capturing part of sequential, dependency, and contextual features from English dataset. The extraction of opinion words from Chinese tourism scenarios remains to be explored. Different from the existing work, we propose a unified representation model that synergistically exploits multiple representations for Chinese data.

B. Sentiment analysis in travel domain

Previous studies in travel domain mainly included opinion mining, sentiment classification, and recommendation, etc. [14] developed a model for aspect category sentiment analysis and review rating prediction on a Chinese hotel review dataset. [15] performed sentiment analysis on reviews to obtain user preferences, and offered personalized recommendation by measuring the semantic similarity between scenic spots and user preferences. [16] proposed an aspect-level sentiment classification framework which consists of a decision tree-based model to obtain explicit and implicit aspects from travel reviews and a classifier to filter important features. [17] tackled event-aware multimodal review sentiment analysis task. They extracted features from different modalities such as text and images, then performed cross-modal association modeling to learn discriminative representations, finally adopted a multi-task learning framework to simultaneously perceive the event type and user sentiment polarity. [18] applied analytical association method for opinion mining on three major travel agencies in China, and demonstrated the usefulness of thematic words, topics and network structural properties for capturing diversity in users concerns. [19] developed a high-quality summarization model that considers both review helpfulness and hotel features. [20] proposed a personalized travel review summarization method, which uses a user-aware sequence network to incorporate users' preference or writing style into the process of summarization generation. [21] presented a controllable aspect-level opinion summarization system, where tourists can set personalized attributes such as summary length and interest of aspects.

Most of the existing works under travel domain focus on aspect and opinion term extraction and sentiment classification. Little research has been done for identifying opinions for a specific aspect which is beneficial for business applications. This motivated us to construct a dataset for TOWE under tourism domain and develop a corresponding target-oriented opinion extraction model.

III. Proposed TRansformer-based Multi-level Model

A. Problem definition and formalizations

Given a travel review $W = (w_1, w_2, ..., w_n)$ and an aspect $A = (w_i, w_{i+1}, ..., w_{i+p})$, where n and p is the length of the review and the aspect, w_i represents a single character in the sentence, the goal is to use the review text W, domain knowledge D such as part-of-speech tags (pos-tag), and task knowledge T such as relative positions to construct a knowledge-enhanced sequence $X = \{w_i, d_i, t_i\}_1^n$, for input model M, extract the fine-grained opinion words $O = (w_j, w_{j+1}, ..., w_{j+q})$, where q is the length of the opinion phrase.

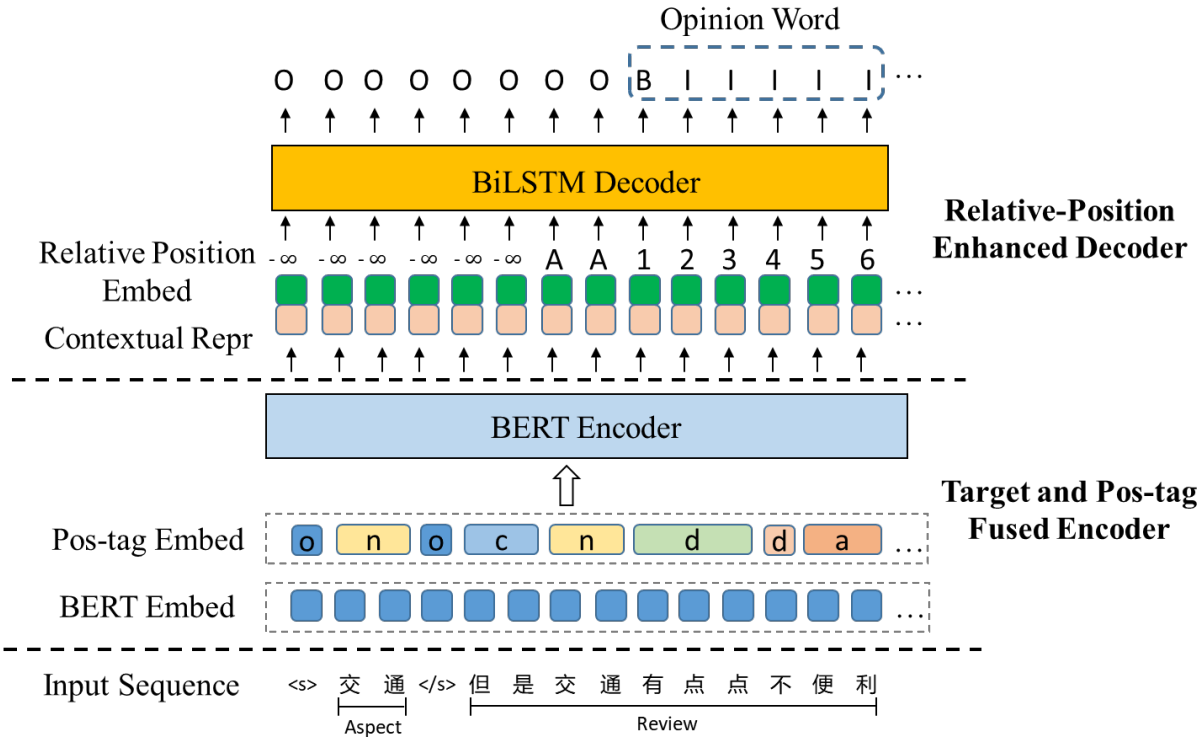

Fig. 2 The overview of the proposed model for target-oriented opinions extraction

B. *System architecture of the proposed model*

Fig. 2 presents a detailed framework of the proposed model. It mainly consists of two modules, namely *Target and Pos-tag Fused Encoder* and *Relative-Position Enhanced Decoder*. In the encoder part, the target word is used as a query to retrieve corresponding opinion from the travel review. We concatenate the target word and the original review in consistence with the input format of the model. Additional linguistic features such as Chinese pos-tags serve as auxiliary information to exploit the dependency between words, and is input together with the concatenated sequence. The target, review and pos-tags labels get fully interacted through multi-layer transformers to generate a target and knowledge-fused contextual representation. To boost the performance, we employ the pretrained BERT as initialization. In the decoder part, a BiLSTM is leveraged to incorporate contextual representation with relative positions to capture position-aware information between aspect and opinion. Finally, a CRF is adopted for sequence labeling.

2.1 Target and pos-tag fused encoder

Target word is a key component in TOWE which is associated with the extracted opinion word. Previous work such as TSMSA[13] adds special token [SEP] surrounding target words and uses multi-head self-attention (BERT) to obtain target-specified representation, since an aspect may appear several times in a review, the method can't identify appropriate insertion position. The machine reading comprehension framework always have the question ahead of the content and can encode task-aware knowledge to facilitate extraction. Therefore, we devise a mechanism to include target-specified information via the formulation like a machine reading comprehension [22, 23]. Specifically, the aspect is employed as a query and preprocessed with the review into the form of $\{[\text{CLS}], a_1, a_2, ...a_p, [\text{SEP}], s_1, s_2, ..., s_n\}$, where [CLS] and [SEP] are special tokens. In addition, SentiLare[1] has shown that linguistic features including part-of-speech tag is closely related to aspect-based sentiment analysis. Under Chinese scenario, pos-tag serve both as a word segmenter and word dependency resolution, which helps extract the complete opinion with degree adverb. Therefore, the Chinese pos-tags (noun(n), verb(v), adverb(d), adjacent(a), Chinese idiom(i) and others(o)) are also included in the input. The text sequence with target and pos-tag is now defined as $X = \{x_i = (w_i, pos_i)\}_1^{n+p}$, where w_i, pos_i means the $i-th$ character and its corresponding part-of-speech label. $n+p$ is the length of the input sequence. Then, X is passed to a multi-layer transformer module to synergistically model target, linguistic feature with vanilla review and obtain a contextual representation $H = \{H^1, H^2, ..., H^l\}$, where l denotes the layers of the transformer blocks. H^i can be calculated with the following formula:

$$O^i = \text{MultiHead}\left(H^{i-1}, m\right) \quad (1)$$

$$FFN^i = \max\left(0, O^i W_1^i + b_1^i\right) W_2^i + b_2^i \quad (2)$$

$$H^i = LN\left(H^{i-1} + FFN^i\right) \quad (3)$$

$$H^0 = E_{char}(\{w\}_1^{n+p}) + E_{seg}(\{i\}_1^{n+p}) + E_{position}(\{i\}_1^{n+p}) + E_{postag}(\{pos\text{-}tag_i\}_1^{n+p}) \quad (4)$$

MultiHead represents multi-head self-attention layer, m is the number of attention heads, W_1^i and W_2^i are transition matrixes, LN is layer-norm operation. The output of the encoder is H^l . E_{char} , E_{seg} , $E_{position}$, E_{postag} denotes character embeddings, segment embeddings(*0* for aspect and *1* for review), absolute positional embeddings and pos-tag embeddings, respectively. Unlike using GNN structure to model the syntactic dependency [11, 12], we employ transformer as fully connected graph neural networks to learn the pos-tag information more effectively.

2.2 Relative-position enhanced decoder

In the decoding part, we perform sequence labeling to find the target-specified opinion. Traditionally, a linear or CRF layer could handle this task. However, it can be seen from Fig. 1 that relative distance between the target and the corresponding opinion can be considered as an effective feature. The transformer is good at modeling global context while recurrent neural network can capture local sequential features. Thus, we enhance the representation learning process via a relative position embedding learning procedure, integrated by a BiLSTM. In detail, the relative position p_i for the $i-th$ character is defined as follows:

$$p_i = \begin{cases} a_{start} - i, \ if \ a_{start} - i \in [1, s^*] \\ i - a_{end}, \ if \ i - a_{end} \in [1, s^*] \\ 0, \ i \in [a_{start}, a_{end}] \\ C, \ else \end{cases} \quad (5)$$

where a_{start} , a_{end} are the start and end index of the assigned aspect, respectively, under Chinese scenario the aspect usually contains more than one character so a_{start} and a_{end} are different. p_i will be assigned as a constant C that is normally set to the maximum relative distance which is denoted by s^* . Note that the distance is calculated based on the nearest mention in the reviews. Then, we adopt BiLSTM to learn the representation h_i from the encoder and relative position embedding pe_i to model the context in a forward and backward direction:

$$r_i = \text{BiLSTM}([h_i; pe_i]) \quad (6)$$

Using relative position to enhance the representation can prevent the target and the opinion from overlapping, and enable extracting more complete opinion like the combination of degree adverbs and adjectives owing to short-term dependency mechanism. Employing the BiLSTM helps pass the relative information through the context.

Finally, Conditional Random Field (CRF) decoding policy is chosen to guarantee the correlations between tags in neighborhoods (B-Opinion, I-Opinion, O) based on the learned r_i . Specifically, we use a linear-chain CRF and score the tag sequence as conditional probability:

$$p(y|\ r) = \frac{\exp(s(r, y))}{\sum_{y' \in Y} \exp\left(s\left(r, y'\right)\right)} \quad (7)$$

where Y is the set of all possible tag sequences and $s(r, y)) = \sum_i^n \left(A_{y_{i-1}, y_i} + P_{i, y_i} \right)$ is the score function, A_{y_{i-1}, y_i} is the transition matrix and P_{i, y_i} is emission matrix.

2.3 Learning Objective

Since we use CRF to decode the sequence, the negative log likehood is defined as the loss and the learning objective is to minimize the negative log probabilities:

$$L(s) = -\log p(y|\ r) \quad (8)$$

IV. Experimental Results and Discussions

A. Data collection and annotation

To validate the performance of the proposed model, we collect a travel dataset from December 1, 2015 to December 1, 2020 from www.dianping.com, which is a famous travel review site in China. We select 16 top-rated tourist attractions in Beijing including the Temple of Heaven, Badaling Great Wall and etc., and invite two annotators to label the target aspect and their corresponding opinion words, the annotators had sufficient research experience on scenic spot review data, the Kappa coefficient of the final annotation result was 0.86, indicating a high agreement. Note that the whole opinion including adverbs of degree is annotated, this setting can be used to analyze the fine-grained sentiment polarity of the tourists. The annotation process adopts the BIO method, "B" denotes the first word of the opinion word, "I" denotes the middle or last word of the opinion word, and "O" denotes the words without opinion. The final dataset consists of 4,663 travel review sentences and 9,488 aspect-opinion pairs. The statistics of the dataset is shown in TABLE I.

TABLE I Details of the Dataset

Item	Count
review count	4663
aspect-opinion pair count	9488
average review length	45.81
max review length	527
min review length	3
average aspect-opinion distance	4.96
max aspect-opinion distance	54
min aspect-opinion distance	1

B. Baselines

To demonstrate the efficacy of the proposed model. We compare various models in TOWE, including Distance-rule, Pipeline, TC-LSTM, IOG, TSMSA.

- Distance-Rule [5]: It uses pos-tagging tool [24] to acquire pos tag for each word and select the nearest adjective from the target as the candidate opinion word.
- Pipeline: This procedure first extracts all the opinion words in the review using a trained LSTM; then selects the closest opinion words of the target as the result.
- TC-LSTM [9]: This method concatenates the averaged target vector with each word in sentence and then feeds them into a BiLSTM for sequence labeling.

- IOG [9] : It utilizes an Inward-Outward LSTM and a Global LSTM to capture the information of aspects and global information, respectively, then combines these information for extraction.
- TSMSA [13]: The target words are extended with [SEP] token, then the review is fed into BERT to generate a target-aware representation for sequence labeling.

C. Evaluation metrics

Precision(P), Recall(R) and F1 score (all in %) are adopted to evaluate the performance, which are calculated by the following formula:

$$precision = \frac{\#correct\_opinion}{\#predicted\_opinion} (9)$$

$$recall = \frac{\#correct\_opinion}{\#annotated\_opinion} (10)$$

$$f1 = 2 * \frac{precision * recall}{precision + recall} (11)$$

where $\#correct\_opinion$ is the number of target-oriented opinions that correctly extracted by the model, $\#predicted\_opinion$ is the number of all the predicted target-oriented opinions, $\#annotated\_opinion$ is the number of the true annotated opinions with the assigned aspect.

D. Implementation details

Following previous work [9], the constructed dataset is randomly divided into the training set, the validation set and testing set according to the ratio of 7:1:2. Hyper-parameters are finetuned on the training set and the best model are saved based on the performance on the validation set. We use Adam optimizer for all the deep neural network models. Models that use LSTM as encoder including Pipeline, Target-LSTM and IOG use pretrained Chinese word embedding from [25] and have a learning rate of 1e-4 with batch size 8. Models that use BERT as encoder including TSMSA and our proposed model use pretrained BERT-Base-Chinese from [26] and have a learning rate of 2e-5 with batch size 4.

E. Experimental results

TABLE II shows the performance comparison of all models. The neural networks-based methods outperform rule-based method significantly. Among the neural networks-based baseline methods, precision varies from 61.74 to 77.21, recall varies from 56.76 to 77.04, whereas F1 varies from 61.14 to 77.12. The pipeline method performs the worst in terms of precision and F1, indicating that the error cascade caused by opinion words extraction may be one of the reasons for limiting the performance of the model. Target-LSTM performs better than IOG, indicating that the fusion mechanism of IOG is better under the condition of provided aspect information. TSMSA performs the best on precision, recall, and F1 among all the baselines, which are 77.21, 77.04, and 77.12, respectively, this is probably due to the better contextual modeling ability of the multi-layer transformer. By further integrating domain and task-aware knowledge such as aspect, part-of-speech and relative position in a hierarchical mode, the proposed method achieves an improvement over TSMSA, with precision of 79.88, recall of 80.23, and F1 of 80.01. This further implies the deep task-aware semantic representation can be learned by combining these information, which plays important roles in fully characterizing the relationship between aspect and opinions.

TABLE II PERFORMANCE OF ALL MODELS

Models	*P*	*R*	*F1*
Distance-Rule	15.66	16.74	16.18
Pipeline	61.74	60.55	61.14
Target-LSTM	67.05	56.76	61.47
IOG	73.33	67.88	70.5
TSMSA	77.21	77.04	77.12
Ours	**79.88**	**80.23**	**80.01**

[a.] best results are marked in bold

F. Analysis on auxiliary features

To study the impacts of auxiliary features, we perform ablation test and show the results in TABLE III.

It can be seen from TABLE III that both pos-tag and relative position contribute to the improvement over model with only aspect information. Compared with pos-tag, the relative position plays a more important role on this task, this might be because relative position indicates the synergy of target and opinion while pos-tag is a kind of word-level information. In the proposed model, except for both the auxiliary information, we use transformers to model the global context of words and word dependency and use BiLSTM with relative position embedding to model the local sequential context between aspect and opinion. When all of them are tackled in a multi-level manner, the model performs the best, which indicates the effectiveness of the fusion mechanism.

TABLE III CONTRIBUTION OF AUXILIARY FEATURES

Models	*P*	*R*	*F1*
+aspect	77.47	76.99	77.23
+aspect +relative-position	**80.41**	78.77	79.58
+aspect +pos-tag	80.32	77.41	78.81
Ours	79.88	**80.23**	**80.01**

[b.] best results are marked in bold

G. Analysis on different embeddings and encoders

To evaluate the impacts of different embeddings and encoders, we conduct several experiments by varying the above factors. We present the comprehensive analysis of the encoder part in ***Variant of the Encoder*** and the decoder part in ***Variant of the Decoder*** in TABLE IV.

In ***Variant of the Encoder***, "random initialized" and "BERT embedding" mean the model use randomly initialized embedding or BERT embedding. Both of them are not initialized with the encoder weight of BERT. "BERT fixed" means we freeze the BERT weight and do not finetuned in the training phase. "BERT finetuned" is exactly the same as the proposed model. The results indicate that he embeddings or the encoder is important for the performance

of the models. Firstly, the model shows poor performance without the pre-trained encoder with/without the pretrained BERT embeddings. Such results show that BERT embedding needs to cooperate with the pre-trained encoder to perform better on TOWE. Secondly, applying the weight of BERT for initialization without fine-tuning also fails on this task. On the one hand, the pretrained corpus of BERT and the travel review have a language gap, so finetuning can help capture the domain-specified aspect. On the other hand, since the auxiliary pos-tag embedding is included as the input, the attention weight calculation needs modifications, implying that the BERT weight needs incremental update.

In the ***Variant of the Decoder*** part, we study the impact of different relative position embedding. "w/o position embedding" denotes the relative position information does not appear in the decoding procedure, while "learned embedding" and "cosine embedding"[27] represent relative position embedding in parametric and nonparametric form, respectively. The detailed implementation of cosine embedding is shown as follows:

$$pe_{pos,2i} = \sin(pos / 10000^{2i/d_{model}}) \quad (12)$$

$$pe_{(pos,2i+1)} = \cos(pos / 10000^{2i/d_{model}}) \quad (13)$$

where *pos* is the position and *i* is the dimension. The experimental results show the importance of the relative position embeddings and the embedding mode.

TABLE IV PERFORMANCE ON VARIANTS OF EMBEDDINGS AND ENCODERS

Models	*P*	*R*	*F1*
Variant of the Encoder			
randomly initialized	67.57	67.66	67.48
BERT embedding	67.66	67.48	67.57
BERT fixed	59.24	60.51	59.24
BERT finetuned	**79.88**	**80.23**	**80.01**
Variant of the Decoder			
w/o Embedding	80.32	77.41	78.81
learned Embedding	**79.88**	**80.23**	**80.01**
cosine Embedding	78.55	81.31	79.9

[c.] best results are marked in bold

H. Case study

Fig. 3 shows an example of extracted target-oriented opinion from travel reviews. The constructed prototype system includes functions such as travel review collection, preprocessing, aspect setting, aspect-oriented opinion extraction and opinion tracing, which allows users to read reviews quickly and effectively. Various aspects such as architecture, accommodation, tickets, service, environment, etc. are identified. Users can explore the mined aspects and keep track of the opinions. Take a tourist that plans to visit ancient architectural cultural attractions in Beijing for an example, due to the busy schedule, he can only choose one scenic spot that satisfies his expectation. By using the system, the user can quickly find the right information about the attractions. When he clicks the button of "Architecture", the system automatically extracts aspect-opinion pairs about the architecture, the representative opinions include "very antique", "very spectacular", "preserved well", "very distinctive", "really worth a visit", "a little bit luxurious". We can observe that degree adverbs and adjectives are all extracted together as opinion words. Assume that the user is interested in "Architecture: Very Antique", he can choose this item and the system will display all the original sentences containing these opinions. By reading the related reviews, he finds that "Beijing Water Town" is a good choice to visit. It can be seen from the above example that the mined aspect-opinion pairs provide users with a more fine-grained understanding of attractions, which is valuable for making appropriate decisions.

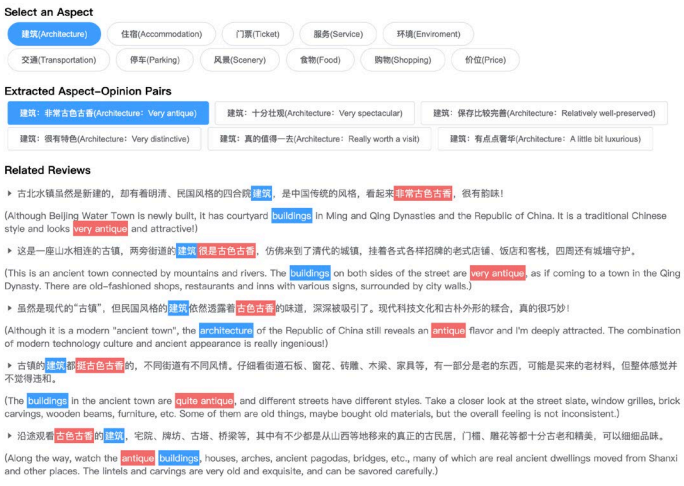

Fig. 3 An example of extracted target-oriented opinion from travel reviews

V. CONCLUSION AND FUTURE WORK

In this paper, we perform target-oriented opinion extraction for Chinese travel review. A multi-level transformer-based fusion model is proposed, which takes advantages of BERT and task-aware knowledge. Experimental results on a constructed Chinese travel review dataset demonstrate the efficacy of the proposed model. In the future, we plan to expand our dataset and perform other tasks such as implicit aspect and opinion extraction [28], Aspect Sentiment Quad Prediction (ASQP) [29] task, etc.

VI. ACKNOWLEDGEMENT

This work was partially supported by the National Key Research and Development Program of China (Grant No.2020AAA0103405), the National Natural Science Foundation of China (Grant No.62071467, 71621002), and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA27030100).

REFERENCES

[1] P. Ke, H. Ji, S. Liu, X. Zhu, and M. Huang, "SentiLARE: Sentiment-Aware Language Representation Learning with Linguistic Knowledge," in *EMNLP*, 2020.

[2] J. Wang, K. Wei, M. Radfar, W. Zhang, and C. Chung, "Encoding Syntactic Knowledge in Transformer Encoder for Intent Detection and Slot Filling," in *AAAI*, 2021.

[3] Y. Bowen *et al.*, "Joint Extraction of Entities and Relations Based on a Novel Decomposition Strategy," *ArXiv,* vol. abs/1909.04273, 2020.

[4] S. K. Sahu, F. Christopoulou, M. Miwa, and S. Ananiadou, "Inter-sentence Relation Extraction with Document-level Graph Convolutional Neural Network," *ArXiv,* vol. abs/1906.04684, 2019.

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews," *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,* 2004.
[6] L. Shu, H. Xu, and B. Liu, "Lifelong Learning CRF for Supervised Aspect Extraction," in *ACL*, 2017.
[7] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Double Embeddings and CNN-based Sequence Labeling for Aspect Extraction," in *ACL*, 2018.
[8] X. Li and W. Lam, "Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction," in *EMNLP*, 2017.
[9] Z. Fan, Z. Wu, X. Dai, S. Huang, and J. Chen, "Target-oriented Opinion Words Extraction with Target-fused Neural Sequence Labeling," in *NAACL*, 2019.
[10] Z. Wu, F. Zhao, X. Dai, S. Huang, and J. Chen, "Latent Opinions Transfer Network for Target-Oriented Opinion Words Extraction," *ArXiv,* vol. abs/2001.01989, 2020.
[11] J. Jiang, A. Wang, and A. Aizawa, "Attention-based Relational Graph Convolutional Network for Target-Oriented Opinion Words Extraction," in *EACL*, 2021.
[12] S. Mensah, K. Sun, and N. Aletras, "An Empirical Study on Leveraging Position Embeddings for Target-oriented Opinion Words Extraction," *ArXiv,* vol. abs/2109.01238, 2021.
[13] Y. Feng, Y. Rao, Y. Tang, N. Wang, and H. Liu, "Target-specified Sequence Labeling with Multi-head Self-attention for Target-oriented Opinion Words Extraction," in *NAACL*, 2021.
[14] J. Bu *et al.*, "ASAP: A Chinese Review Dataset Towards Aspect Category Sentiment Analysis and Rating Prediction," in *NAACL*, 2021.
[15] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Syst. Appl.,* vol. 167, p. 114324, 2021.
[16] M. Afzaal, M. Usman, and A. C. M. Fong, "Tourism Mobile App With Aspect-Based Sentiment Classification Framework for Tourist Reviews," *IEEE Transactions on Consumer Electronics,* vol. 65, pp. 233-242, 2019.
[17] L. Wang, W. Guo, X. Yao, Y. Zhang, and J. Yang, "Multimodal Event-Aware Network for Sentiment Analysis in Tourism," *IEEE MultiMedia,* vol. 28, pp. 49-58, 2021.
[18] Z. Hou, F. Cui, Y. Meng, T.-h. Lian, and C. Yu, "Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis," *Tourism Management,* 2019.
[19] C.-F. Tsai, K. Chen, Y.-H. Hu, and W.-K. Chen, "Improving text summarization of online hotel reviews with review helpfulness and sentiment," *Tourism Management,* vol. 80, p. 104122, 2020.
[20] J. Li, H. Li, and C. Zong, "Towards Personalized Review Summarization via User-Aware Sequence Network," in *AAAI*, 2019.
[21] R. Mukherjee, H. C. Peruri, U. Vishnu, P. Goyal, S. Bhattacharya, and N. Ganguly, "Read what you need: Controllable Aspect-based Opinion Summarization of Tourist Reviews," *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval,* 2020.
[22] X. Li, J. Feng, Y. Meng, Q. Han, F. Wu, and J. Li, "A Unified MRC Framework for Named Entity Recognition," *ArXiv,* vol. abs/1910.11476, 2020.
[23] J. Liu, Y. Chen, K. Liu, W. Bi, and X. Liu, "Event Extraction as Machine Reading Comprehension," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.
[24] W. Che, Y. Feng, L. Qin, and T. Liu, "N-LTP: A Open-source Neural Chinese Language Technology Platform with Pretrained Models," *ArXiv,* vol. abs/2009.11616, 2020.
[25] S. Li, Z. Zhao, R. Hu, W. Li, T. Liu, and X. Du, "Analogical Reasoning on Chinese Morphological and Semantic Relations," in *ACL*, 2018.
[26] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *NAACL-HLT*, 2019.
[27] A. Vaswani *et al.*, "Attention is All you Need," *ArXiv,* vol. abs/1706.03762, 2017.
[28] H. Cai, R. Xia, and J. Yu, "Aspect-Category-Opinion-Sentiment Quadruple Extraction with Implicit Aspects and Opinions," in *ACL/IJCNLP*, 2021.
[29] W. Zhang, Y. Deng, X. Li, Y. Yuan, L. Bing, and W. Lam, "Aspect Sentiment Quad Prediction as Paraphrase Generation," in *EMNLP*, 2021.