Research on Opinion Targets Extraction of Travel Reviews Based on RoBERTa Embedded BILSTM-CRF Model

Zeyu Li *
State Key Laboratory of Media Convergence and Communication
Communication University of China
Beijing, China
lizeyu@cuc.edu.cn

Tianhe Yu
State Key Laboratory of Media Convergence and Communication
Communication University of China
Beijing, China
cuc_yutianhe@cuc.edu.cn

Hao Shen
State Key Laboratory of Media Convergence and Communication
Communication University of China
Beijing, China
shenhao@cuc.edu.cn

***Abstract*—Tourism is a form of cultural expression. The key step of fine-grained sentiment analysis of travel review evaluation text is opinion targets extraction. It aims to analyze travel review text and extract the opinion targets contained in it, and targets extraction's accuracy rate directly affects the accuracy rate of sentiment analysis. Due to the outstanding performance of pre-training language models in the field of natural language processing in recent years, in order to improve the accuracy of opinion targets extraction, we propose a RoBERTa pre-training language model and Chinese pre-training word embedding vector, combined with BiLSTM (bidirectional long short-term memory) and CRF (conditional random field) opinion targets extraction model. The experimental results on the travel review data set of the Mafengwo travel software show that the extraction effect of this model is improved to varying degrees compared with other existing opinion targets extraction models based on deep learning.**

Keywords—RoBERTa , opinion targets extraction，BiLSTM，CRF

I. INTRODUCTION

Tourism is a form of culture, and the development of tourism can effectively expand the radiation channels of cultural resources. With the rapid development of the Internet and the widespread popularity of mobile devices such as mobile phones, travel software has become an important part of today's travel ecology. Therefore, travel software represented by Mafengwo and Ctrip has developed rapidly, and the number of user comment texts in the software has also increased exponentially in recent years. Tourists are the main body of tourism consumption, and their evaluation of scenic spots and itineraries has important reference value for scenic area managers to improve scenic spot facilities and services. In this context, object-level fine-grained sentiment analysis has become one of the hot research fields. The key step in object-level fine-grained sentiment analysis is opinion targets extraction.

The opinion targets refers to the subject who bears the emotional attitude in a certain comment, which is specifically expressed as the object modified by the evaluation words in the comment text[1], such as a certain scenic spot, event, person, etc. evaluated in travel reviews. Previous opinion targets extraction mostly used traditional machine learning methods and rule-based methods. The former is represented by CRF (conditional random field), which treats the extraction of opinion targets as a sequence labeling task for processing[2],and the latter is committed to building better rule templates to extract opinion targets through matching methods[3], however, the accuracy and adaptability of these extraction methods need to be improved. With the wide application of deep learning in the field of natural language processing, extraction methods based on deep learning have begun to be widely studied in recent years.

As the experiment of Wu Xiaoli et al. [7] proved that short text is vectorized by word vectors, BiLSTM (bidirectional long short-term memory) is used to extract the context features of short texts, and the attention mechanism is introduced to dynamically adjust the importance of features. The experiments on two types of data sets prove that the words the vector representation method is based on the effectiveness of Self-Attention and BiLSTM algorithms in short text sentiment classification. Therefore, this paper proposes a RoBERTa-BiLSTM-CRF opinion targets extraction model based on the RoBERTa pre-training model [16] and Chinese pre-training words embedding vector, combined with BiLSTM and CRF. The experimental results confirm the effectiveness and feasibility of this method.

II. RESEARCH STATUS AND BACKGROUND KNOWLEDGE

A. *Current status of research in the field of opinion targets extraction*

After more than ten years of research on the extraction of opinion targets, there are already many feasible methods and models. At first, the Apriori algorithm was used to find opinion targets from frequent nouns [18], the principle is to count the word frequency. Essentially, opinion targets extraction is a special case of the information extraction problem. Therefore, Li et al. [19] integrated Skip-CRF and Tree-CRF to extract opinion targets. The characteristics of these two CRFs are the fact that they can not only learn word sequences, but also find structural features. In recent years, with the extensive application of deep learning and neural networks in the field of natural language processing, the accuracy of opinion targets extraction has also been greatly improved. Nguyen et al. [20] added phrases and dependency syntax and other information into the recurrent neural network, and proposed PhraseRNN the model enriches the content of the evaluation information. Chen et al. [21] proposed a dynamic multi-pooling convolutional neural network to automatically extract the features of words and sentences. Zhang et al. [22] combined neural network and CRF, and introduced word embedding to improve the accuracy of extraction. Sun et al. [8] proposed an opinion targets extraction model that combines long and short-term memory network and attention mechanism. This model is also one of the latest models in the field of Chinese text opinion targets extraction. This paper was utilized it as a control group in the experimental part.

B. *Text representation of character granularity*

The opinion targets extraction task generally adopts the sequence labeling method combined with CRF. However, because there is no sharp dividing line between words in Chinese text, the inaccuracy of word segmentation will lead to error transmission of text representation based on word granularity. Research of Li Weikang et al. [9] directly proved that for Chinese text, the representation based on character granularity will be more accurate. However, since Chinese texts often use words to refer to the object of evaluation, it is a waste to completely abandon the boundary information of word segmentation. You can combine the commonly used sequence labeling methods on the basis of character granularity and add the position information of the possible word segmentation of the characters. Improving the extraction effect. Common sequence labeling methods include BIO, BIEO, etc. In order to improve the efficiency of labeling, this paper adopts the BIO labeling method, B represents the beginning character of the opinion targets (Begin), I represents the internal character of the opinion targets (In), and O represents the non-opinion targets character (Out). An example of initial text annotation based on character granularity is shown in Table 1:

TABLE I. INITIAL ANNOTATION EXAMPLE

Dataset	Opinion targets	Annotate results
Mafengwo Tourism Evaluation Dataset	玛瑙寺	午/O,后/O,阳/O,光/O,下/O,的/O,玛/B,瑙/I,寺/I,安/O,静/O,美/O,丽/O,。/O

C. *Attemtion*

The attention mechanism was first applied to the field of image processing. In 2015, Bahdanau et al. [10] first introduced the attention mechanism to machine translation. Now this mechanism has become a research hotspot in the field of natural language processing. The core idea of the attention mechanism is to allocate limited attention resources by calculating the importance of the input content, essentially calculating the degree of association between the source data and the target data. Yang Shanliang et al. [11] defined a weight formula to quantify this degree of association, the calculation formula is as follows:

$$\alpha_t = \frac{\exp(f(m_t,m_s))}{\sum_{s\in S}\exp(f(m_t,m_s))} \tag{1}$$

where ms refers to the source data, mt refers to the target data, and the normalized denominator is the sum of all source data and target data. The value obtained from formula (1) is normalized with the softmax function, and what is obtained is the probability distribution of the source data on the corresponding target data. The formula for the function $f(m_t, m_s)$is as follows:

$$f(m_t,m_s) = \begin{cases} m_t^T m_s & dot \\ v_a^T \tanh(W_a m_t + U_a m_s) & concat \\ m_t^T W_a m_s & bilinearity \\ W_a[m_t; m_s] & scaling \end{cases} \tag{2}$$

The Self-Attention mechanism is a special attention mechanism, also called intra attention, whose source data is the same as the target data. When using the Self-Attention mechanism to extract opinion targets, this paper calculates the dependency between each character in the input sentence and other characters, and the calculation process is the same as the attention mechanism.

D. *Recurrent Neural Network*

Recurrent Neural Network (RNN) was suggested by the 1980s.Long short-term memory network (LSTM) is a particular type of recurrent neural network. It uses a delicate structure called "gate" to control the flow of information in and out of neuron cells. It creatively solves the inherent "gradient" in traditional recurrent neural networks. The problems of "explosion" and "gradient disappearance" also avoid the problems of long-term dependence, so the LSTM network is very suitable for processing and forecasting data based on time series [12]. The three main stages inside LSTM include the forgetting stage, the selective memory stage and the output stage are realized by forgetting gate, input gate and output gate respectively. The specific structure is shown in Figure 1:

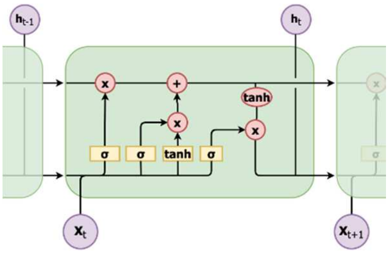

Fig. 1. LSTM unit structure

Suppose the input vector at time t is X_t, the state of the LSTM neural unit is c_t, the output of the hidden layer is h_t the state of the neuron at time t-1 is c_{t-1}, and the output of the hidden layer is h_{t-1} .Then the control formula of the "forgotten gate" at time t is:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X_t] + b_f) \tag{3}$$

where W_f is the weight matrix of the forget gate; b_f is the bias term of the forget gate; $[h_{t-1}, X_t]$ indicates the splicing of two vectors; σ indicates the activation function; f_t indicates that it should be retained How much is the state of the unit at t-1.

Similarly, the input gate control formula at time t is:

$$i_t = \sigma(W_i \cdot [h_{t-1}, X_t] + b_i) \tag{4}$$

where W_i is the weight matrix of the input gate; b_i is the input gate bias term; $[h_{t-1}, X_t]$ represents the splicing of two vectors; σ represents the activation function. The input of the unit state at time t is jointly determined by the output at time t-1 and the input at time t. The formula is as follows:

$$c_t' = \tanh(W_c \cdot [h_{t-1}, X_t] + b_c) \tag{5}$$

where W_c is the weight matrix; b_c is the bias term;

The unit state ct at time t is determined by the unit state ct-1 at time t-1 and the calculation results of formulas (3), (4), and (5). The formula is as following:

$$c_t = f_t \odot c_{t-1} + i_t \odot c_t' \tag{6}$$

where ⊙ represents the multiplication of elements in the corresponding position of the matrix

Then the control formula of the "output gate" at time t is:

$$o_t = \sigma(W_o \cdot [h_{t-1}, X_t] + b_o) \quad (7)$$

where W_o is the weight matrix; b_o is the bias term;

The final output of LSTM at time t is obtained by multiplying the unit state and the judgment condition obtained by the output gate, the formula is as follows:

$$\mathrm{h_t} = o_t \cdot \tanh(c_t) \quad (8)$$

III. The opinion targets extraction model of this article

This paper proposes a RoBERTa-BiLSTM-CRF opinion targets extraction model based on Chinese pre-trained word embedding vectors, and its model structure is shown in Figure 2:

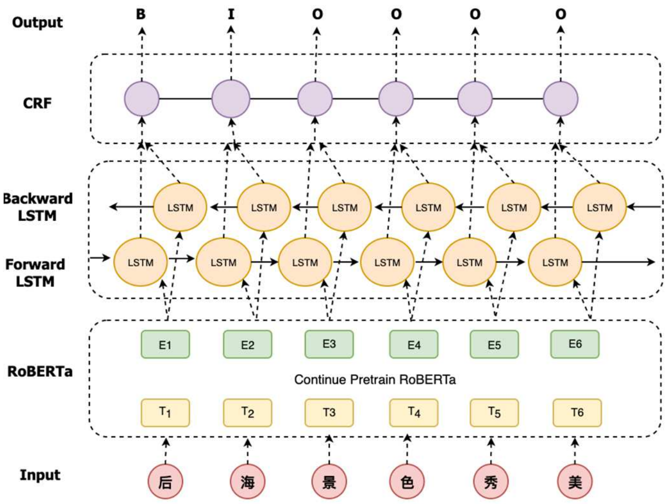

Fig. 2. RoBERTa-BiLSTM-CRF model structure diagram

A. Model hierarchy

The model proposed in this paper is divided into five layers, which are input layer, RoBERTa pre-training layer, BiLSTM layer, CRF layer, and output layer. The core of them is the inner three layers. The working principles and methods of these three layers will be introduced in detail below.

B. RoBERTa pre-trainin

For the task of evaluating targets extraction, data is at the heart part. Therefore, the quality of data preprocessing and pre-training directly influences the accuracy of targets extraction. In 2018, Devlin J et al. [13] proposed a BERT language model based on character vectors and Attention mechanism. Compared with the aforementioned popular ELMo model [14], it further strengthened the generalization ability and is better suitable for sentence level and word Natural la-nguage processing tasks such as level and character level. Output coding vector of BERT is the unit sum of three embedding features [15].

The three embedding features are Token Embedding, Position Embedding and Segment Emb-edding, as showed in Figure 3:

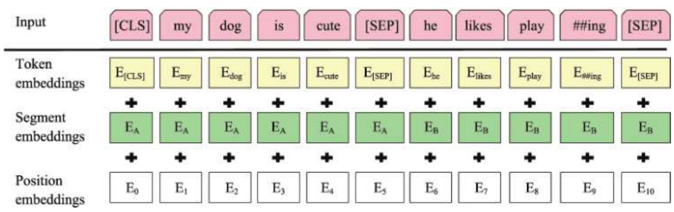

Fig. 3. BERT input vector representation

The BERT model has two self-supervised tasks, namely Masked Language Model (MLM) and Next Sentence Prediction (NSP). The former refers to randomly masking some words from the input corpus during training, and then telling the word through the context, similar to cloze. The latter is tantamount to predict whether there is a contextual relationship between two sentences. Compared to other language models, BERT has stronger contextual long-distance semantic learning capabilities. However, Liu Y et al. [16] found that there is still abundant room for improvement after in-depth research on the BERT model. Liu Y et al. used a longer time, a larger batch size, and more data for training. While removing the NSP target in the BERT, they also dynamically changed the mask mode according to the training data, and finally got the improvement of the BERT model. The model is the RoBERTa pre-training model used throughout this article.

C. BiLSTM

Due to its design characteristics, LSTM is very suitable for modeling text data. However, LSTM can only retain past information, and cannot understand the contextual features of Chinese characters. Therefore, this paper uses BiLSTM as showed in Figure 4, which essentially constructs two LSTM neural networks in opposite directions at the same time, which can better capture the two-way semantic dependence. Among them, historical feature information of each character can be obtained through the forward LSTM, and the following feature information of the character can be obtained through the reverse of LSTM, and finally the context-related word vector can be obtained.

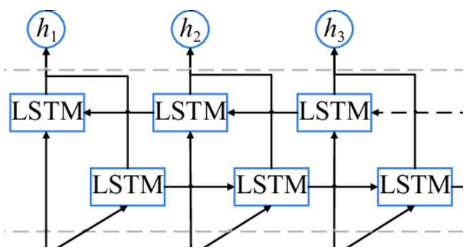

Fig. 4. BiLSTM

The hidden layer of BiLSTM also saves the forward output vector $\vec{h}$ and the reverse output vector $\overleftarrow{h}$. Then the output of the module at time t is:

$$\mathrm{h_t} = \overrightarrow{h_t} + \overleftarrow{h_t} \quad (9)$$

D. CRF

CRF is an undirected graph model proposed by Lafferty J et al.[17],which tends to deal with sequence labeling problems in the field of natural language processing. Input a given set of variable sequences, the model can calculate and output the probability distribution of the corresponding sequence.In the extraction model proposed in this paper, the vector sequence of length n output by the BiLSTM layer is also the input sequence of the CRF layer, then the input observation sequence of the CRF can be set as $X = \{x_1, x_2, x_3, \cdots, x_n\}$, the output label prediction sequence is $y = \{y_1, y_2, y_3, \cdots, y_n\}$. Let Y(X) be the output sequence corresponding to the input sequence, and the marking of each label sequence is $S(X, y)$, then the relevant calculation formula is as follows:

$$S(X, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_i, y_i \quad (10)$$

$$P(y|X) = \frac{\exp(S(X,y))}{\sum_{y' \in Y(X)} \exp(S(X,y'))} \quad (11)$$

$$y_{max} = \text{argmax}\{P(y|X|), y \in Y(X)\} \quad (12)$$

where y_{max} is the labeling sequence with the highest probability.

IV. Experiment and Analysis

A. Experimental data and experimental environment

The data set used throughout this experiment is the comment text crawled on the Mafengwo APP. After cleaning, filtering and batch labeling, there are a total of 58,934, which are all related to tourism. In this paper, 41,253 of them are used as training data, 5,893 are invoked as verification data, and 11,788 are used as test data.The experimental environment of the experiment in this article is shown in the table:

TABLE II. LAB ENVIRONMENT

Hardware environment	CPU：Intel I7-9700 3.00GHz
	RAM: 16.0GB
Software Environment	Windows 10 64bit Python 3.6.0
	tensorflow 1.14.0 keras 2.4.3 gensim 3.4.3 jieba 0.42.1

B. Evaluation criteria

In the evaluation experiment of this article, evaluation standards adopted are international common standards, including precision rate, recall rate, and F_1 value. Their definitions are as following:

$$\text{precision} = \frac{\text{Number of opinion targets correctly extracted}}{\text{Number of all opinion targets extracted}}$$

$$\text{recall} = \frac{\text{Number of opinion targets correctly extracted}}{\text{Number of all opinion targets}}$$

F_1 value is quite special, essentially the harmonic mean of the precision rate and the recall rate, and its calculation formula is (13)~(15):

$$P = \frac{TP}{TP+FP} \quad (13)$$

$$R = \frac{TP}{TP+FN} \quad (14)$$

$$F_1 = \frac{2*P*R}{P+R} \quad (15)$$

P and R in formulas (13) and (14) are precision rate and recall rate, respectively. The meanings of TP, FN, and FP in the formulas are shown in Table 3:

TABLE III. CONFUSION MATRIX

	Positive	Negative
True	True Positive(TP)	True Negative (TN)
False	False Positive(FP)	False Negative (FN)

C. Experimental design and comparative experiment

Training of the RoBERTa-BiLSTM-CRF model uses fixed RoBERTa parameters and only updates the targets extraction method of BiLSTM-CRF parameters. In order to prove the effectiveness of the model in this paper, three sets of controlled experiments were set up, namely:

BiLSTM-CRF: This method adds a CRF layer to the BiLSTM model. In check to see the opinion targets extraction effect without data pre-training, this paper selects it as the comparison method.

Word2Vec-BiLSTM-CRF: This method uses the Word2Vec word vector model to pre-train the data, and adds the BiLSTM layer and CRF layer. In order to verify that the extraction effect based on character granularity is better than the effect of word granularity, this article selects it as a comparison method.

BERT-BiLSTM-CRF: This model is an extraction model based on character granularity, and it is also one of the best opinion targets extraction models at present. In check to see the degree of optimization of the RoBERTa model compared to the BERT model, this article selects it as a comparison method.

D. Experimental parameter settings

The RoBERTa model and the BERT model used in this article are both Google's open source Chinese pre-training models, both with 12 layers, the hidden layer is 768 dimensions, and the 12-head model, while the Word2Vec Chinese model is trained through the open source news corpus of Sogou Lab , Its feature vector is 400 dimensions, and the rest of the parameters are the default values in genesis. The parameters of the opinion targets extraction model in the training process are shown in Table 4:

TABLE IV. MODEL TRAINING PARAMETER

Model	epochs	Sequence_length	batch_size
BiLSTM-CRF	80	——	512
Word2Vec-BiLSTM-CRF	80	50/70/100	512
BERT-BiLSTM-CRF	80	50/70/100	512
RoBERTa-BiLSTM-CRF	80	50/70/100	512

E. Analysis of experimental resultss

This paper has conducted experiments on 4 models under the condition that Sequence_length is 50, 70, and 100. The final results are shown in the following three tables:

TABLE V. PERFORMANCE COMPARISON WHEN SEQUENCE_LENGTH=50

Model	precision	recall	F_1
BiLSTM-CRF	0.8425	0.7357	0.7854
Word2Vec-BiLSTM-CRF	0.8547	0.7716	0.8122
BERT-BiLSTM-CRF	0.8896	0.8304	0.8590
RoBERTa-BiLSTM-CRF	**0.9006**	**0.8463**	**0.8726**

TABLE VI. PERFORMANCE COMPARISON WHEN SEQUENCE_LENGTH=70

Model	precision	recall	F_1
BiLSTM-CRF	0.8425	0.7357	0.7854
Word2Vec-BiLSTM-CRF	0.8584	0.7813	0.8180
BERT-BiLSTM-CRF	0.8926	0.8385	0.8647
RoBERTa-BiLSTM-CRF	**0.9054**	**0.8476**	**0.8755**

TABLE VII. PERFORMANCE COMPARISON WHEN SEQUENCE_LENGTH=100

Model	precision	recall	F_1
BiLSTM-CRF	0.8425	0.7357	0.7854
Word2Vec-BiLSTM-CRF	0.8672	0.7730	0.8174
BERT-BiLSTM-CRF	0.8972	0.8392	0.8672
RoBERTa-BiLSTM-CRF	**0.9122**	**0.8468**	**0.8783**

The results shown in Table V, Table VI, and Table VII the precision, recall and F_1 of the RoBERTa-BiLSTM-CRF model are all the best.On the whole, the above models have achieved good results when Sequence_length=70, and the performance after Sequence_length is increased to 100 is basically the same as before, which also shows that the setting of the sequence length parameter has a certain effect on the final extraction result influences. From the comparison of indicators, the precision of each model is generally higher than the recall rate.

From the perspective of the granularity of the model, the opinion targets extraction effect of the character vector-based extraction model is significantly better than that of the word vector-based model. Moreover, whether it is the Word2Vec-BiLSTM-CRF extraction model based on word vectors or the RoBERTa-BiLSTM-CRF model and the BERT-BiLSTM-CRF model based on character vectors, their extraction effects are considerably better than the BiLSTM-CRF extraction without pre-training model.

From the exact data point of view, the model proposed in th-is paper achieves the best extraction effect proposed in this paper when Sequence_length=100. The F1 value of the model in this paper is up to 12% higher than BiLSTM-CRF, and it is also high-er than BERT-BiLSTM-CRF, which is nearly 1.3% .

In summary, the opinion targets extraction model of tourism reviews proposed in this paper has a good effect.

V. CONCLUSION

In order to further improve the accuracy of opinion targets, this paper proposes a travel review opinion targets extraction model based on a pre-trained language model. First, the input text is converted into a character vector through the RoBERTa pre-training model, and the two-way long and short-term memory network is used to Recognize the context of each character, transform the opinion targets extraction task into a sequence labeling task, and finally use the CRF model to find the optimal labeling sequence, so as to accurately identify the opinion targets. Related experiments conducted in this paper also strongly prove that the method in this paper is effective.

Future research can consider combining the RoBERTa pre-training model with other mainstream neural networks, and comparing the performance of new model combinations, such as the performance comparison of models such as RoBERTa-BIGRU-CRF, RoBERTa-CNN-LSTM-CRF, etc., can also be considered Explore the best parameter combinations and best application scenarios of each model in this article.

VI. REFERENCES

[1] Liu, and Bing. "Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies." Synthesis Lectures on Human Language Technologies 5.1(2012):1-167.

[2] Jakob, N. , and I. Gurevych . "Extracting Opinion Targets in a Single- and Cross-Domain Setting with Conditional Random Fields." Conference on Empirical Methods in Natural Language Processing 2010.

[3] Song Xiaolei, Wang Suge, and Li Hongxia . "Research on Automatic Recognition of Product Evaluation Object for Specific Domain." Journal of Chinese Information 24.1(2010):89-94.

[4] A, Soujanya Poria , E. C. B , and A. G. C . "Aspect extraction for opinion mining with a deep convolutional neural network." Knowledge-Based Systems 108(2016):42-49.

[5] Irsoy, O. , and C. Cardie . "Opinion Mining with Deep Recurrent Neural Networks." Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2014.

[6] Zhang Pu, etc. "Deep Learning Based Method of Chinese Microblog Evaluation Object Extraction ." Computer Engineering and Design 39.08(2018):246-250+301.

[7] Wu Xiaohua , etc. " Sentiment Analysis of Chinese Short Text Based on Self-Attention and BI-LSTM." Journal of Chinese Information 033.006(2019):100-107.

[8] Sun Haizhou. "Research on the extraction of evaluation objects of news reviews." 2020. University of Chinese Academy of Sciences (School of Artificial Intelligence, Chinese Academy of Sciences), MA thesis.

[9] Li Weikang, Li Wei, and Wu Yunfang. "Exploration of the Combination of Chinese Character Vectors and Word Vectors in Deep Learning." Journal of Chinese Information Processing 31.6(2017).

[10] Bahdanau, D. , K. Cho , and Y. Bengio . "Neural Machine Translation by Jointly Learning to Align and Translate." Computer Science (2014).Yang Shanliang, and Sun Qi. "Recurrent Neural Network Evaluation Object Extraction Model Based on Attention Mechanism." Computer Applications and Software 36.03(2019):208-215.

[11] Yang Shan and Sun Qi. "Recurrent Neural Network Evaluation Object Model Based on Attention Research." Computer Applications and Software 36.3 (2019): 202-209.

[12] Tang Kang, Wang Haitao, and Jiang Ying. "Research on E-commerce Review Evaluation Object Extraction Based on Deep Learning." Electronic Measurement Technology 13(2019).

[13] Devlin, J. , et al. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." (2018).

[14] Peters, Matthew , et al. "Deep Contextualized Word Representations." Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers) 2018.

[15] Wu, Y. , et al. "Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation." (2016).

[16] Liu, Y. , et al. "RoBERTa: A Robustly Optimized BERT Pretraining Approach." (2019).

[17] Lafferty, J. D. , A. K. Mccallum , and F. Pereira . "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." (2001).

[18] Popescu, A. M. , and O. Etzioni . "Extracting product features and opinions from reviews." (2005).

[19] Li, F. , et al. "Structure-Aware Review Mining and Summarization." International Conference on Computational Linguistics Association for Computational Linguistics, 2010.

[20] Nguyen, T. H. , and K. Shirai . "PhraseRNN: Phrase Recursive Neural Network for Aspect-based Sentiment Analysis." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing 2015.

[21] Chen, Y. , et al. "Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks." The 53rd Annual Meeting of the Association for Computational Linguistics(ACL2015) 2015.

[22] Zhang, Meishan, Yue Zhang, and Duy-Tin Vo. "Neural networks for open domain targeted sentiment." Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015.