Aspect Based Sentiment Analysis of catering field reviews via RoBERTa-AOA model

Bo Li1, Fang Pan2*, Zhaoyu Shou3 and Huibing Zhang1

1 Guangxi Key Lab of Trusted Software, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China
2 Teachers college for vocational and technical education, Guangxi Normal University, Guilin, Guangxi, 541004, China
3 School of Information and Communication, Guilin University of Electronic Technology, Guilin, Guangxi, 541004, China

*Corresponding author’s e-mail: panfang@mailbox.gxnu.edu.cn

Abstract. The emotional information included in catering field reviews has a great influence on the ordering behaviour of users. The traditional Aspect Sentiment Analysis Model based on Neural Network often has the problem of neglecting Aspect Words. This paper proposes an aspect sentiment analysis model that combines RoBERTa and Attention-over-Attention (AOA). Firstly, the model uses RoBERTa to model sentences and aspects respectively. Then, the deep semantic features of sentence and aspect were extracted by Bi-LSTM. Finally, AOA is adopted to learn the weight of aspect to sentence and sentence to aspect respectively, and the results are input into SoftMax layer for classified output. Experimental results on real Meituan catering data sets show that the accuracy of RoBERTa-AOA is about 1.7%-3.7% higher than that of other models.

1. Introduction

The development of the Internet has changed people's consumption and behaviour habits. More and more consumers leave a large number of consumption reviews on online platforms. Theses review texts contain a large amount of emotional information. How to fully excavate and analysis users' emotional tendencies from the catering review information has a great impact on improving the quality of catering service.

At present, catering sentiment analysis mainly includes three methods: sentiment dictionary, machine learning and deep learning. For example, Vishnu et al. used review data in multiple fields to build an emotional dictionary adapted to a specific field based on multi-field knowledge updates [1]. The performance of the emotional dictionary-based method is inseparable from rules and dictionary construction, but the construction of rules and dictionary requires professional knowledge and a lot of manpower. The machine learning method needs to strictly process the data, and then extract the features after extraction, which takes a lot of time, and classification is performed after the above steps are completed. For example, Kiritchenko et al. combined emotion dictionary and SVM to extract aspect level emotional polarity from SemEval2014 data set and achieved good results [2]. However, model training relies on the construction and selection of high-quality features of annotated data sets, which requires a lot of work costs.
Different from emotion dictionary and machine learning method, deep learning emotion analysis can learn hidden information of data and extract it automatically, avoiding manual feature extraction, and alleviating to some extent the problems that the relationship between texts is difficult to measure and highly dependent on artificial construction features. For example, Hassan et al. proposed to remove the pooling layer that is easy to lose information in CNN, and replace LSTM with long-distance information, which can reduce the loss of local exhaustive information and effectively learn long-distance dependent information between sentences [3]. Dong et al. proposed a model combining multiple functions with Ada-RNN to solve the problem of judging the emotion expressed by different aspects of sentences [4]. Liang Bin et al. raise a multi-attention convolutional neural network method for sentiment analysis of specific targets. This method greatly reduces the time of model training. By combining various attention methods, deeper emotional feature information can be obtained, and the emotional polarity of different targets can be effectively identified on the data set [5].

Researchers have used the attention mechanism combined with deep learning to train the data, with good results in the field of sentiment analysis. For example, Zhou et al. propose a way to use Bi-LSTM network based on attention to capture important semantic information in sentences [6]. Tang et al. based on the ability of LSTM to obtain a better dependent information set, and used it to perform well in aspect sentiment analysis, and thus proposed a TD-LSTM model. By changing the combination structure of input context and aspect words on the basis of TD-LSTM model, a TC-LSTM model is proposed. This model input aspect words and context by means of stitching, so as to make aspect words and context more closely and establish a better connection [7].

In the above models of sentiment analysis, the importance of aspect words is considered, but the independent modelling of aspect words is neglected. The paper proposes an aspect sentiment analysis model that combines RoBERTa and Attention-over-Attention (AOA). In this model, RoBERTa pre-training model is used to dynamically code aspect words and sentences, generate word vectors of context-dependent information respectively, and input the obtained word vectors into the Bi-LSTM model respectively. Use the model of Bi-LSTM can capture information which rely on bidirectional semantic, then learn the important parts of aspect and context through AOA. After obtaining the final representation, it is input to the classification layer to classify the emotion polarity corresponding to different aspects information contained in the catering review. Experiments were conducted on the fine-grained user reviews sentiment analysis data set for the catering field in the 2018 Global AI Challenge. The experimental results indicate that the RoBERTa-AOA model put forward in this paper can attain higher accuracy compared with other sentiment analysis models.

2. Problem definition and model

2.1. Problem definition
When analysis the emotions of the review of Catering field, this paper defines sentences \( s = \{w_1, w_2, w_3, \cdots, w_n\} \), aspect \( a = \{w_1, w_2, w_3, \cdots, w_m\} \). The aspect contained in the review text can be composed of phrases or a single word. The result purpose of the model is to be able to identify the emotional polarity of a certain aspect \( a \) in the catering review sentence \( s \).

The overall structure proposed in this paper is shown in Figure 1. The RoBERTa-AOA model is mainly composed of the following four parts: word embedding layer, semantic representation layer, cross attention layer and classification output layer.

2.2. Word embedding layer
Using RoBERTa to pre-train the word embedding layer of sentences and aspects, you can get a better representation of sentences and aspects. RoBERTa uses Transformer as the main framework of the algorithm, which can more thoroughly capture the two-way relationship of sentences.

The RoBERTa pre-training model maps each word \( w \) of the catering review sentence and aspect to a low-dimensional vector \( v_w \in \mathbb{R}^{d_w} \), where \( d_w \) is the word vector dimension. The output of pre-training
sentences and aspects is the catering review sentence vector \( \{v_1^s, v_2^s, v_3^s, \cdots, v_n^s\} \in R^{n \times d_v} \) and the aspect vector \( \{v_1^a, v_2^a, v_3^a, \cdots, v_m^a\} \in R^{m \times d_v} \).

![Figure 1. Overall model structure](image)

2.3. Semantic representation layer
After the sentence and aspect are encoded with word vectors, they are input into the semantic representation layer respectively. In order to solve the long-distance problem, the semantic representation layer combines the LSTM in the forward and backward directions. LSTM has a good effect on capturing long-distance dependent information of sentences, and effectively avoids the problem of gradient disappearance or explosion, but there is a problem that the dependent information can only be extracted in one direction. For example: "The restaurant's service attitude is not good, not as good as the next door", the "no" here is a modification of the "service attitude". Bi-LSTM can effectively capture the bidirectional semantic relationship that LSTM cannot obtain. The calculation formula of Bi-LSTM is shown in formula (1) and formula (2)

\[
\overrightarrow{h}_t = LSTM([v_1^s, v_2^s, v_3^s, \cdots, v_n^s]) \tag{1}
\]

\[
\overleftarrow{h}_t = LSTM([v_1^s, v_2^s, v_3^s, \cdots, v_n^s]) \tag{2}
\]

\[
h_t = [\overrightarrow{h}_t, \overleftarrow{h}_t] \tag{3}
\]

Among them, the input is represented as \( s = [v_1^s, v_2^s, v_3^s, \cdots, v_n^s] \), \( \overrightarrow{h}_s \), \( \overleftarrow{h}_s \) are the forward and backward hidden sequence state, and finally the final output \( h_s \) is obtained by splicing \( \overrightarrow{h}_s \) and \( \overleftarrow{h}_s \). In this way, the sentence and the aspect are calculated to obtain the sentence and aspect respectively. The hidden state of \( h_s \in R^{n \times 2d_v} \) and \( h_a \in R^{m \times 2d_v} \), \( d_v \) is the dimension of the hidden state.

2.4. AOA layer
AOA layer takes the aspect \( h_a \in R^{m \times 2d_v} \) and sentence \( h_s \in R^{n \times 2d_v} \) obtained by the semantic representation layer to calculate the influence weight of the aspect and sentence respectively through AOA. First, calculate the matching matrix \( I = h_s \cdot h_a^T \); then, perform SoftMax on the rows and columns of the matching matrix \( I \) to get the attention degree \( \alpha \) of the sentence to the aspect and the attention degree \( \beta \) of the sentence to the aspect respectively; finally, calculate the average of the attention \( \beta \).
from the sentence to the aspect, Get aspect-level attention $\beta \in \mathbb{R}^w$, $\bar{\beta}$ is used to represent the most important part of the aspect. The aspect-level attention $\bar{\beta}$ and the attention degree $\alpha$ are weighted to calculate the final impact $\gamma \in \mathbb{R}^w$. As shown in the following formula (7):

$$
\alpha_i = \exp(I_i) / \sum_i (I_i) \\
\beta_i = \exp(I_i) / \sum_i (I_i) \\
\bar{\beta}_i = \frac{1}{n} \sum_r (\beta_r) \\
\gamma = \alpha \cdot \bar{\beta}
$$

2.5. Classification output layer

The classification output layer carries out dot product sum between the results of AOA layer and the results of semantic representation layer, and inputs them to the classification output layer to predict the final emotional polarity corresponding to aspects in catering field reviews. As shown in formula (9):

$$
r = h^T \cdot \gamma \\
p = \text{softmax}(w^T r + b)
$$

Among them, $p \in \mathbb{R}^C$ is the probability distribution of emotional polarity in catering reviews. Among them, $W$ and $b$ represent weight matrices and bias matrices respectively in formula (9). $C = 3$ represents the three types of emotional polarities corresponding to catering reviews.

The model is trained through cross-entropy loss function and $L2$ regularization. As shown in formula (10):

$$
L = -\sum_i \sum_{c \in C} [(y_i = c) \cdot \log(p(y_i = c)) + \lambda \| \theta \|_2^2]
$$

Among them, $y_i = C$ is the true emotional polarity in catering reviews, $p(y_i = c)$ is the predicted emotional polarity in catering reviews, and $\lambda$ is a regularization parameter.

3. Experiment

3.1. Data set

This paper uses the 2018AI Global Competition for the fine-grained user reviews sentiment analysis data set in the catering field. The experimental data statistics are shown in Table 1.

|           | Positive | Neutral | Negative |
|-----------|----------|---------|----------|
| **Train** | 365739   | 135499  | 49754    |
| **Test**  | 19895    | 7029    | 3481     |

3.2. Parameter settings

The parameter settings of the RoBERTa-AOA model are shown in Table 2.
3.3. Experimental results and analysis

3.3.1. Experimental evaluation index. This page uses accuracy and F1-score as measurement indicators for sentiment analysis and comparative experimental models of catering review text.

| Parameter name         | Parameter value |
|------------------------|-----------------|
| Vector dimension       | 768             |
| Maximum sequence length| 300             |
| Batch size             | 16              |
| Learning rate          | 1e-5            |
| Epoch                  | 4               |
| Dropout                | 0.1             |
| Optimizer              | Adam            |

3.3.2. Analysis of results. For further verification the performance of the model, Use the following models to compare with the model described in this paper:

- **ATAE-LSTM**: The extension of the AT-LSTM model combines sentences and aspect words into the input of the model by appending aspect words to each word vector.
- **RoBERTa**: Input the sentence and the aspect word as two sentences, separate the sentence and the aspect word with the [SEP] symbol and predict the relationship between the two sentences.
- **RoBERTa-BiLSTM**: Sentences and aspect words are used as the input of RoBERTa, and then deep semantic features are obtained through Bi-LSTM.
- **Word2vec-BiLSTM-AOA**: Use word2vec to model the sentence separately, input the word vector obtained into the Bi-LSTM to get the hidden layer state, and focus on the important part of the sentence through AOA.

| Model                  | Accuracy | F1-score |
|------------------------|----------|----------|
| ATAE-LSTM              | 0.688    | 0.490    |
| RoBERTa                | 0.815    | 0.741    |
| RoBERTa-BiLSTM         | 0.831    | 0.769    |
| Word2vec-BiLSTM-AOA    | 0.835    | 0.776    |
| RoBERTa-BiLSTM-AOA     | 0.852    | 0.796    |

From Table 3, it is easy to see the accuracy and F1 corresponding to different models. The RoBERTa-AOA proposed in this paper reaches 85.2%, which is 1.7% and 2.1% higher than Word2vec-BiLSTM-AOA and RoBERTa-BiLSTM respectively. The RoBERTa pre-training model can encode word vectors well, and the cross-attention mechanism can learn the influence weights of sentences and aspect words, so the accuracy is higher.

Among the selected baseline models, ATAE-LSTM without an attention mechanism performed the worst. Compared with ATAE-LSTM, Word2vec-BiLSTM-AOA with the attention mechanism added to the data set has an accuracy rate of about 15%. The attention mechanism can be used to learn important parts of sentences. In addition, the effect of RoBERTa-BiLSTM is better than that of Word2vec-BiLSTM-AOA. The reason is that RoBERTa has a self-attention mechanism Transformer,
which can learn the relationship between context and aspect words very well, while Bi-LSTM can extract long-distance dependence Relationship, the extracted semantics are more accurate.

4. Conclusion
In the aspect sentiment analysis task in catering field, the paper proposes an aspect sentiment analysis model fusing RoBERTa and AOA. First, RoBERTa models sentence and aspect respectively, and the obtained sentence and aspect word vectors are input to the Bi-LSTM model to catch the adjective information of the hidden state of the sentence and aspect, and finally the important parts of the sentence and aspect are automatically focused through the AOA model. Compared with other traditional methods, the experimental results of the model proposed in this paper verify its effectiveness. Since the maximum input length of RoBERTa is 512 words, it has certain limitations for long text comment texts exceeding 512 words. In the following study and work, we will continue to explore the difficulty of processing long text by adapting the input length to the Roberta model by truncating long text.

Acknowledgments
This work is supported by National Natural Science Foundation (61662013, 61967005, U1811264); Guangxi Innovation-Driven Development Project (Science and Technology Major Project) (AA17202024); The Guangxi Natural Science Foundation (2017GXNSFAA198372).

References
[1] Vishnu, K. S., Apoorva, T., Gupta, D. (2014) Learning domain-specific and domain-independent opinion oriented lexicons using multiple domain knowledge. In: 2014 Seventh International Conference on Contemporary Computing IEEE. pp. 318-323.
[2] Kiritchenko, S., Zhu, X., Cherry, C., et al. (2014) NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews. In: Proceedings of the 8th International Workshop on Semantic Evaluation, SemEval2014. Dublin, Ireland.
[3] Hassan, A., Mahmood, A. (2017) Deep Learning approach for sentiment analysis of short texts. In: International Conference on Control. pp. 705–710.
[4] Dong, L., Wei, F., Tan, C., et al. (2014) Adaptive Recursive Neural Network for Target-dependent Twitter Sentiment Classification. In: Meeting of the Association for Computational Linguistics. pp 49–54.
[5] Liang, B., Liu, Q., Xu, J., Zhou, Q. (2017) Target-specific Sentiment Analysis Based on Multi-attention Convolutional Neural Network. Computer Research and Development. 1724-1735.
[6] Zhou, P., Shi, W., Tian, J., et al. (2016) Attention-Based Bidirectional Long Short-Term Memory Networks for Relation Classification. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. pp207.
[7] Tang, D., Qin, B., Feng, X., Liu, T. (2015) Effective LSTMs for target-dependent sentiment classification. Computer. Science. 3298–3307.