# ABSA for Chinese review with fine-tuned BERT and zeroshot classification

## **Zichen Wang**

School of Computer Science & Engineering, South China University of Technology, Guangzhou, China

#### 202030134018@mail.scut.edu.cn

**Abstract.** In recent years, there has been a trend towards utilizing Large Language Models (LLMs) to address diverse issues in computational linguistics. More parameters and layers make them better suited to approximating the objective function in multi-domain tasks. In the twenty-first century, since e-commerce has grown quickly, people's interest in gleaning insightful information from other people's evaluations has grown. Conducting fine-grained sentiment analysis is becoming more beneficial. Aspect-Based Sentiment Analysis (ABSA) is a sub-problem within the field of Sentiment Analysis. The ABSA problem encompasses two sub-problems: Identification of targets and Opinion Mining. Existing methods address ABSA well with a relationship between the aspect and the labels. It is the most common situation in application combined and fine-tuned Bidirectional Encoder Representations from Transformers (BERT) model. The method is used experimentally in five tags (1-5) at the "service" and "environment" levels. In the experiment, the predictions of the method-derived model are contrasted with those of the original model. The results show a substantial improvement in the model's prediction accuracy in different aspects.

Keywords: Bert, ABSA (Aspect Based Sentiment Analysis), Zero-shot Classification.

## 1. Introduction

In both academics and industry, sentiment analysis is seen as a crucial endeavor. However, most current algorithms ignore the aforementioned entities and their features in favor of attempting to determine the overall polarity of a phrase, paragraph, or text range. The goal of aspect-based sentiment analysis (ABSA) is to more thoroughly mine comments [1]. People's remarks in daily life have several facets. The automobile, for instance, this car has a lovely cabin but a bad driving experience. The significance of the *interior* and *driving feel* is unknown, hence there is no use in dissecting the mood of this statement at a crude level. However, when reading a review, consumers frequently focus on the review's interesting passages and ignore its unimportant ones.

Previously, the difficulties in ABSA reviews of Arabic hotels are discussed, along with cuttingedge methods based on supervised machine learning [2]. Ma et al. proposed the use of attentive LSTM to solve the challenges of ABSA and Targeted SA using common sense knowledge. [3]. To explicitly extract aspect and opinion words, Wang et al. suggested a unique joint model that unifies recurrent neural networks with conditional random fields [4]. The suggested model simultaneously transfers information across aspect and opinion words while learning high-level differentiating characteristics. Gating mechanisms and convolutional neural networks form the basis of the model that Xue et al. claim is the most accurate and efficient method [5]. Xu et al. proposed a Review Reading Comprehension (RRC) problem intended to solve the problem of how to convert user comments into a large amount of valid information for machine terms. Simultaneously, Xu et al. recommended a new dataset named ReviewRC that can be used to fine-tune pre-trained models for the RRC problem [6]. To use the contextual reading comprehension capabilities of the BERT model to solve ABSA problems, Sun et al. proposed a training method utilizing sentence pairs to enhance the comprehension of aspects and sentiments [7]. ABSA research has gradually changed from finding more accurate syntactic and semantic feature representations to gaining experience through deep learning. The structure of the network using machine learning ABSA went from simple to complex with a large increase in parameters.

Pre-trained models, such as Bidirectional Encoder Representations from Transformers (BERT) [8], have more recently provided novel solutions to the ABSA problem. All that is necessary to achieve the desired outcomes at the application level is to add an output layer to the model and then modify its internal parameters using a specific dataset. Azzouza et al. recommended four steps for Twitter Sentiment Analysis [9]. Zhou et al. provide several pre-training exercises using emoticons, sentiment words, and ratings that are free of human input and are at both the token and sentence levels [10]. Huang et al. use a pre-trained language model to perform entity-level sentiment analysis on financial information, predicting market direction based on the emotional polarity of articles on financial websites [11]. Therefore, it is obvious that LLMs have been widely used in various fields with good results.

Given the excellent performance of LLMs in various subtasks of computational linguistics, it is expected to fine-tune the pre-trained model using a Chinese review dataset so that it can meet the needs of language understanding and evaluation of Chinese sentiment in various aspects.

Firstly, this paper proposes an ABSA method based on zero-shot classification combined with finetuning BERT and finally proves the method's effectiveness on the data set without the strong relations between labels and specific sentences. This paper also completes an ABSA task based on a Chinese comment dataset.

# 2. Related work

In this chapter, work related to this paper will be mentioned. The pre-trained models on which the experiments are based will likewise be presented.

## 2.1. Transformer

This article's two pre-trained models were created using the transformer proposed by Vaswani et al. [12]. The transformer's structure consists of three parts: encoder, decoder, and output. The encoder receives input from the entire model, while the decoder receives information from past training. In contrast to the encoder, which has a multi-head attention unit and a feed-forward neural (FFN) network, the decoder has an extra multi-head attention unit with a mask, which can well handle chronological tasks.

Transformer was first used for machine translation tasks. What led to its generalization to various LLMs is the high parallelism and reduced training time using large amounts of data. The Transformer's high degree of parallelism comes from the matrix computation in the attention mechanism. The computation of the attention mechanism involves calculating the similarity of two vectors k and q representing key and query. Assuming a k matrix K, a q matrix Q, and a v matrix V (value), the similarity matrix S is computed as eq.1 shows:

$$\mathbf{S} = \mathbf{Q} \times \mathbf{K}^{\mathrm{T}} \tag{1}$$

W is the weight matrix of the output values:

$$S_{i}^{'} = \frac{e^{s_{i,j}}}{\sum_{i} e^{s_{i,j}}},$$
(2)

$$\mathbf{W}_{i} = \mathbf{S}_{i}^{'} \cdot \mathbf{V}_{i}^{\mathrm{T}} \,. \tag{3}$$

To address the issue of excessive self-focus, a multiple attention mechanism is proposed. The transformer model uses the method of independently learning to obtain h sets of different linear projections to transform the K, Q, and V matrices, with the projection matrices as  $W^K, W^Q, W^V$ . Then, these h sets of transformed K, Q, and V matrices are pooled in parallel for attention. Finally, the outputs of these h attentional poolings are spliced together and an inner product is done by another linear projection matrix  $W^0$  that can be learned to produce the final output. Matrix computations can be utilized in place of the conventional linear method in this fashion. The time-consuming in training the model can be reduced by processing matrix calculations in parallel.

## 2.2. BART and zero-shot learning

Bidirectional and Auto-Regressive Transformers (BART) (https://huggingface.co/facebook/bart-largemnli) is selected as a model as a tool for entity extraction evaluation since the dataset lacks entity labels, and zero-shot classification is required [13]. Lampert et al. first proposed zero-shot learning to solve challenging data labeling in computer vision [14]. Zero-shot learning can also be applied in the field of computational linguistics. The BART model for text zero-shot learning is due to its superior performance and ease of use [15]. Consequently, the ABSA problem is broken down into a problem of sentence-level sentiment analysis.

## 2.3. BERT

BERT (https://huggingface.co/bert-base-chinese) has been extensively utilized across various subfields in computational linguistics and demonstrated excellent performance in text classification assignments. Sun et al. presented specific techniques to fine-tune BERT and proved that the fine-tuned BERT model can attain high accuracy in text classification assignments [16]. As shown in Figure 1, the text already includes instances of relevant entities and their corresponding labels (labels with values 1 to 5), rendering the sentiment analysis issue into a text multi-classification task.



Figure 1. Structure of aspect-based sentiment analysis workflow.

# 3. Experiment

The pre-processing of the data utilized in experiments is covered in this chapter. The process of model training and loss function calculation is presented together with the experimental model and its fine-tuning. Figure 1 depicts the whole aspect-based sentiment analysis procedure. First, the comment data is divided into aspect *environment* and aspect *service* categories. The BERT model and the fully connected layer are then iteratively trained using these comments in two aspects.

# 3.1. Dataset and zero-shot classification

Yf\_dianping is a Chinese review dataset with different aspects and scores (1-5). Additionally, certain data sets lack comments or have one or more aspect ratings that are blank. During the data cleaning procedure, these data are removed.

The dataset has four ratings (1-5) with user comments. The four ratings include an overall score and three specific aspect scores. Of the three aspect scores, the *flavor* is ambiguous in Chinese and is not included in the sentiment classification experiment but in the zero-shot classification. There are two other aspect scores, *service*, and *environment*, which do not conflict with each other and are clear to be found in texts, so they are both targeted aspects of experimental data for fine-tuning BERT.

**Table 1.** Example of comments after zero-shot classification (translated from Chinese).

Comments	Environment	Service
It is bigger and environmentally better than Carrefour, and even better than Goodwill, but some products are not as cheap as Goodwill. [Environment] It's the best supermarket I've been to in Chengdu and has a huge advantage over Carrefour, Goodwill, and Red Flag. The waiters are very nice and helpful. The staff are very nice and helpful, and they sufficiently do things, which makes you feel comfortable shopping. [Service]	4	4
The cake price is not expensive, over the birthday in his house to buy a cake of what coffee, not very good, cake and do not side high. And when you buy a cake, the waiter is not caring. [Service]	3	2
There was no one there when I went there and it was a great setting. [Environment] ???? I ordered beef tenderloin with foie gras gnocchi, but what came up was a whole beef. When I asked the waiter, he said that he had changed the menu for me temporarily and that the steak was some kind of import and cost \$270 a piece, which wasn't on the menu, but I could pay the original price of \$178 for it. If I want to change back, have to do now time will be long, the waiter said very justified, feel or I took advantage of the same, dizzy![Service] ???? I was really hungry at the time and didn't bother with him, so I didn't change it, but the steak was tempting. I'm sure it's good, and the taste of the steak is superb, I want it to be 8 minutes rare, but it's still very tender, it's the best I've ever eaten! ???? The Caesar salad had a very flavourful and very special dressing and we both ate a lot of it after not liking vegetables! ???? The lasagne is also the most unusual I've ever had, it's not as greasy as it is made elsewhere, it feels like it's all freshly ground sauce with fresh tomatoes, and it's delicious! ???? I think it would have been a perfect dinner if there	5	3

The fine-tuning of BERT needs to include the text and its scores, so it requires using zero-shot classification to identify the sentences with corresponding scores in the comments. In canonical Chinese expressions, often only one complete meaning is expressed in a sentence. Therefore, this paper performs sentence-level sentiment analysis on a sentence-by-sentence basis. In addition, the separator between sentences in canonical Chinese expressions is often "o" "!" "? " "......" ";" and so on.

However, it is not excluded that there are irregular expressions in the dataset, such as " "t"," "," "," "," "?" "^" etc. as sentence separators. So these symbols have also been included as sentence delimiters.

After data cleaning and segmentation, BART is used for zero-sample classification. The result obtained is the probability value of the possible labels corresponding to the sentence. The relationship of this probability can be represented as eq.4 shows:

$$P(environment) + P(flavor) + P(service) = 1.$$
 (4)

The threshold for the probability is set to 0.5. Once there is a probability that the value of P(X) > 0.5, the sentence will indisputably belong to aspect X. The result of data preprocessing is presented in Table 1.

## 3.2. Structure connected with the output of BERT

The [CLS] labels from the final hidden layer of the BERT model may be easily retrieved, resulting in a pooled output of a tensor (with a batch size of 8 and embedding dimension of 768). To obtain information that is essential to the labeling operation, a multi-layer perceptron (MLP) must be added to this output. The completely linked layer proposed in this study is shown in Figure 2. The probability of a text's label membership is obtained by applying a Softmax function to the MLP's final outputs. The score label corresponding to the maximum of these probabilities is considered to be the score of the aspect to which the text corresponds. And the tensor composed of these five probabilities will be used to calculate the loss.



Figure 2. The structure of a fully connected layer and output.

## 3.3. Fine-tuning BERT and fully connected layer

The loss function selected is cross entropy and the loss of each epoch is the arithmetic mean of training text. The loss of one text x can be calculated as eq.5 shows:

$$H(x) = \sum_{i=1}^{5} -P(x_i) \log P(x_i).$$
 (5)

The optimizer chosen was AdamW, and the learning rate is 1e-5. There are 50 training epochs overall. The BERT pre-training model's network topology and that of the succeeding fully connected layer's structure do not alter throughout training. Layer by layer, backpropagation changes the values of each network parameter.

Table 2. Comparison with BERT without fine-tuning in Macro F1, Micro F1, and Weighted F1.

	Macro F1		Micro F1		Weighted F1	
	Service	Environment	Service	Environment	Service	Environment
BERT(origin)	0.064	0.070	0.115	0.131	0.042	0.051
BERT(fine-tuned)	0.429	0.359	0.575	0.566	0.541	0.510

## 4. Experiment result

In this chapter, three evaluation indicators for multi-classification missions will be presented to measure the effectiveness of the multi-classification task implementation.

Separate calculations are made to determine the accuracy of the sentiment classification for the environment and the service. Each of these two aspects comprises 5 labels (numbered 1–5). Some labels have both anticipated value i and true value i. The amount of labels have genuine value other and anticipated value i. The number of labels with the true value i and the anticipated value other is  $FN_i$ .

Precision of label *i* is  $P_i$ :

$$P_i = \frac{TP_i}{TP_i + FP_i} \times 100\%.$$
(6)

Recall of label *i* is  $R_i$ :

$$R_i = \frac{TP_i}{TP_i + FN_i} \times 100\%.$$
<sup>(7)</sup>

Macro F1:

MacroF1=
$$\frac{2 \times \sum_{i=1}^{5} R_{i} \times \sum_{i=1}^{5} P_{i}}{5 \times \left(\sum_{i=1}^{5} R_{i} + \sum_{i=1}^{5} P_{i}\right)}.$$
 (8)

Micro F1:

MicroF1=
$$\frac{2 \times \sum_{i=1}^{5} TP_i}{2 \times \sum_{i=1}^{5} TP_i + \sum_{i=1}^{5} FP_i + \sum_{i=1}^{5} FN_i}$$
. (9)

Weighted F1( $w_i$  is the weight of each classification):

Wighted F1 = 
$$\frac{2 \times \sum_{i=1}^{5} (R_i \times w_i) \times \sum_{i=1}^{5} (P_i \times w_i)}{5 \times \left[\sum_{i=1}^{5} (R_i \times w_i) + \sum_{i=1}^{5} (P_i \times w_i)\right]}.$$
(10)

The results of indicators in aspects of *service* and *environment* are shown in Table 2.

# 5. Conclusion

This paper presents an aspect-based sentiment analysis method using zero-shot classification combined with a fine-tuned BERT model. This approach is applied to a dataset of Chinese reviews where text and labels do not contain explicit correspondences. In this application, the method is used experimentally to make a five-category classification for each of the two aspects of the review. Ultimately, the method is shown to be effective. However, the method still needs improvement. When doing zero-shot classification, there is a loss in the classification result, which is difficult to calculate. In addition, the time consumption of calling two pre-trained models is substantial. Also, data labels were uneven during training (scores of 1 and 5 were significantly less than others). This resulted in labels rated 1 and 5 needing to be more accurate than the other scores during testing. Thus, there is still room for optimization, including fine-tuning BART using a dataset containing aspects and their corresponding sentences.

## References

- [1] Maria P, Dimitris G, John P, Harris P, Ion A, and Suresh M 2014 In Proc. of the 8th Int. Workshop on Semantic Evaluation 27-35.
- [2] Al-Smadi M, Qawasmeh O, Al-Ayyoub M, Jararweh Y, and Gupta B 2018 J. Comput. Sci-Neth. 27 386-93.
- [3] Yunkun M, Haiyun P, and Erik C 2018 Proc. of the AAAI Confer. on Art. intel. 32 No. 1.
- [4] Wenya W, Sinno Jialin P, Daniel D, and Xiao X 2018 ArXiv Preprint ArXiv 1603 06679.
- [5] Xue W and Tao L 2018 ArXiv Preprint ArXiv 1805 07043.
- [6] Hu X, Bing L, Lei S and Philip S Y 2019 ArXiv Preprint ArXiv 1904 02232.
- [7] Chi S, Luyao H, and Xipeng Q 2019 ArXiv Preprint ArXiv 1903 09588.
- [8] Jacob D, Ming-Wei C, Kenton L, and Kristina T 2018 ArXiv Preprint ArXiv 1810 04805.
- [9] Noureddine A, Karima A, Roliana I 2020 Spri. Inter. Pub. 428-437.
- [10] Jie Z, Junfeng T, Rui W, Yuanbin W, Wenming X and Liang H 2020 In Proc. of the 28th Intern. Confer. on Comput. Ling. 568-79.
- [11] Zhihong H, Zhijian F 2020 2020 IEEE 18th Inter. Conf. on Indus. Infor. (INDIN). 1.2020.
- [12] Ashish V, Noam S, Niki P, Jakob U, Llion J, Aidan N G, Lukasz K, and Illia P 2017 Advan. in neu. Infor. Proc. Sys. 30.
- [13] Mike L, Yinhan L, Naman G, Marjan G, Abdelrahman M, Omer L, Ves S and Luke Z 2019 ArXiv Preprint ArXiv 1910 13461.
- [14] Christoph H L, Hannes N, Stefan H 2009 In 2009 IEEE Conf. on Comp. Vis. and Patt. Reco. 951-58.
- [15] Ariel G, Alon H, Eyal S, Yotam P, Liat E and Noam S 2022 ArXiv Preprint ArXiv 2210 17541.
- [16] Chi S, Xipeng Q, Yige X and Xuanjing H 2019 18th China National Conf. 18 194-206.