Implicit Aspect Extraction from Online Clothing Reviews with Fine-tuning BERT Algorithm

Li Yu* and Xuefei Bai
Shanghai Key Laboratory of Financial Information Technology, Shanghai University of Finance and Economics, Shanghai, China

*Corresponding author email: yu.li@mail.shufe.edu.cn

Abstract. In the era of e-commerce, tremendous product reviews can provide quick and valuable insight into the market trends. Aspect-based sentiment analysis (ABSA) on product reviews is becoming increasingly important for both companies and consumers. Implicit aspect extraction, a subtask of ABSA, aims to identify the aspect of a review sentence without explicit aspect terms. For Chinese online reviews, this paper proposes a fine-tuning BERT based method to extract implicit aspects. Fine-tuning BERT performs well on a variety of natural language processing tasks. Using labeled data, the hidden layer parameters of pre-training BERT are fine-tuned in a supervised manner to generate fine-tuning BERT. Through experiments on Chinese online reviews of clothing, it is believed that fine-tuning BERT can improve the effect of implicit aspects identification.

Keywords: Online Review; Implicit Aspect; Fine-tuning BERT.

1. Introduction
Nowadays, the huge volume of product reviews tends to be a valuable resource for knowledge extraction. It is believed that online reviews can help companies better understand customer needs, find product defects, improve product quality, and earn more profits [1]. According to the statistics of Xu et al. [2], 30% of Chinese reviews contain implicit aspects. Being an important subtask of aspect-based sentiment analysis (ABSA), implicit aspect extraction is aimed to extract the aspect that does not appear explicitly in the review sentence. Implicit aspect extraction is helpful for companies to understand customer satisfaction towards products more comprehensively. However, it is not an easy job since the dictionary-based methods do not work [3].

In general, there are three methods for implicit aspect extraction. The unsupervised method aims to extract the implicit aspect from the unlabeled corpus. Sun et al. [4] extract implicit aspects through a filtered co-occurrence matrix between explicit aspects and corresponding sentiment words. Karagoz et al. [3] count the number of co-occurrence of a sentiment word and an aspect. Then the sentiment word with the highest count is used to identify implicit aspects. The semi-supervised method aims to extract the implicit aspect from both the labeled and the unlabeled corpus. Xu et al. [2] extract implicit aspects by combining a semi-supervised method of support vector machine (SVM) and topic models. Using K-means algorithm, Hai et al. [5] sort aspects into different clusters to extract implicit aspects. The supervised method aims to extract the implicit aspect from the labeled corpus. Mohammed et al. [6] use a hybrid method of corpus, WordNet, and naive Bayes classifier to extract implicit aspects. Ray and Chakrabarti [7] use a set of rule-based approaches combined convolutional neural network (CNN) approach to conduct aspect extraction.

Since fine-tuning BERT performs well on a variety of natural language processing tasks [8], this paper proposes a fine-tuning BERT based method to extract implicit aspects for Chinese online reviews. This method
method can be divided into three steps. Firstly, data preprocessing such as removing duplicates and labeling are conducted. Secondly, the BERT model is pre-trained by Google on Chinese general corpus. Then, the model is trained with the labeled data to fine-tune the parameters of the hidden layers. Finally, this fine-tuning BERT is used to extract implicit aspects. Experiments show that fine-tuning BERT can improve the effect of implicit aspects extraction, comparing with the pre-training BERT model and the word2vec+LSTM model.

The rest of the paper is organized as follows. Section 2 presents the methodology of implicit aspect extraction with fine-tuning BERT. A case study of clothing reviews and experimental results are presented in Section 3, which is followed by conclusions and extension.

2. Implicit Aspect Extraction Based on Fine-tuning BERT

The framework of the proposed implicit aspect extraction method based on fine-tuning BERT can be seen in Figure 1. The framework can be divided into three parts: data preprocessing, fine-tuning BERT construction, and implicit aspect extraction.

2.1. Data Preprocessing

Data preprocessing contains three subtasks. (1) Online product reviews are segmented to a set of sentences based on delimiters such as space, period or semi-colons. In this paper, each result sentence is assumed to have only one implicit aspect. (2) Duplicate clauses, misspellings, and special symbols are removed. (3) For subsequent supervised learning, the sentences are manually labeled with their implicit aspects. Suppose there are \( N \) sentences, each sentence is represented by \( C_i \), where \( i \in \{1,2,\ldots,N\} \). The jth character in the ith sentence is represented by \( W_{ij} \), where \( j \in \{1,2,\ldots,M_i\} \) and \( M_i \) is the number of characters in the ith sentence. Suppose a product has \( Q \) aspects, each aspect is represented by \( F_i \). Then, the ith sentence is assigned with an aspect \( T_i \), \( T_i \in \{F_1,F_2,\ldots,F_Q\} \). After manually annotation, the product review corpus are stored as two-tuples, namely, \( <C_i,T_i>\).

2.2. Fine-tuning BERT Construction

2.2.1. Five parts of Fine-tuning BERT. The fine-tuning BERT model has five parts, that is, input layer, embedding layer, encoder layer, pooling layer, and softmax layer. In the input layer, every review sentence is padded to the same length since BERT can only take fixed length of sentence as input. Suppose the input length of the BERT model is \( P (P \leq 512) \), and 512 is the maximum input length of BERT. Firstly, for the ith review sentence that contains \( \sum_j W_{ij} \) tokens, the [CLS] token is added at the beginning of the sentence. Secondly, the empty tokens are added to the end of the review sentence until the number of the tokens reaches \( P \). In the embedding layer, every token of a review sentence is processed by summing up three embeddings. Taking the jth character \( W_{ij} \) in the ith sentence as an example, its representation is constructed as a \( n \)-dimensional vector \( TV_{ij} \in \mathbb{R}^d \), combining the token embedding, the segment embedding, and the position embedding. Then, \( TV_{ij} \) is fed to the encoder layer.

In the encoder layer, each \( n \)-dimensional vector \( TV_{ij} \) is the input of a multi-layer bidirectional Transformer encoder. The structure of the encoder is a series of multiple hidden layers with the same
structure but different parameters. It is used to extract the semantic features of the review sentences. The correspond output of the vector $TV_{ij}$ is a vector $OV_{ij} \in \mathbb{R}^H$.

In the pooling layer, the output of the encoder $OV_{ij}$ is linearly transformed to a $h$-dimensional vector $OV_{i0}$ for the [CLS] token of the $i$th review sentence. Since the vector of the [CLS] token is an aggregated sentence-level representation, it is a better expression for text classification tasks [9]. Therefore, the output of the pooling layer is the vector $OV_{i0}$ corresponding to the [CLS] token.

In the softmax layer, each vector $OV_{i0}$ is mapped to a product aspect. This layer acts as a classifier. It is a fully connected layer using softmax function as the activation function. For the $i$th sentence, the output of the softmax layer is represented by $DV_i \in \mathbb{R}^Q$. Each element in the vector $DV_i$ represents the probability of the implicit aspect $a$ ($a \in \{1,2,\ldots,Q\}$) that the $i$th sentence contains. The aspect with the largest probability is considered as the final result of implicit aspect extraction.

2.2.2. Training Process of Fine-tuning BERT. Training process contains two phases: Pre-training and fine-tuning. In the pre-training phase, the first four parts of fine-tuning BERT is conducted for training on a large corpus. There are many types of BERT models pre-trained by Google. This paper uses the BERT model pre-trained on Chinese general corpus. The number of Transformer blocks is 12, the number of self-attention heads is 12, and the hidden size $H$ is 768 [8].

In the fine-tuning phase, the pre-training model is fine-tuned by using the labeled online reviews of clothing. The two-tuples $\langle C_i, T_i \rangle$ generated by data preprocessing is fed into pre-training BERT to fine-tune the parameters of the five parts of fine-tuning BERT in a supervised manner. After the fine-tuning phase, the implicit aspect extraction model for clothing reviews is obtained.

2.3. Implicit Aspect Extraction

In online review corpus, implicit aspect sentences are mixed with explicit aspect sentences. To extract implicit aspect sentences, readers can follow the idea proposed by Karagoz et al. [3]. It can efficiently find sentences without explicit aspect words in the corpus, with a well-defined explicit aspect word dictionary. For an extracted implicit aspect sentences without the label of aspect, it is preprocessed and fed into the implicit aspect extraction model based on fine-tuning BERT. The model will automatically determine the implicit aspect that the review sentence belongs to.

3. Experiments and Results

3.1. Dataset

Clothing reviews often do not contain explicit aspect words. Moreover, some words in Chinese general corpus have different meanings than in clothing reviews. Taking “Mickey” as an example, it generally refers to the name of an animation protagonist in Chinese general corpus, while in clothing reviews, it generally refers to the pattern printed on clothes.

A total of 1150 implicit aspect sentences are randomly selected from the online clothing review corpus provided by a clothing brand company. According to the evaluation index system of clothing e-commerce sellers proposed by Zhu et al. [10] and the word frequency information of the clothing reviews, this paper labels 1150 clothing reviews into 6 aspects: fabric, style, size, color and pattern, price, and others. The last aspect mainly includes consumers’ evaluation of customer service and delivery. In order to fine-tune the model and test the generalization ability of the model, the data is randomly divided into training set and test set, the ratio is 4:1.

3.2. Model Results

The following hyper-parameters are set for the fine-tuning process: the length of input $P$ is set as 50, the number of clothing aspects $Q$ is set as 6, the number of iteration epochs is 5, the learning rate is 1e-5, and the batch size is 8. Then, the test set is used to test the generalization ability of fine-tuning BERT, and the results are shown in Table 1. Table 1 shows that, in six different aspects, the precision, recall, and F1-score are all very large. However, fine-tuning BERT performs poorly on color and pattern aspect, and there are two reasons. On the one hand, the words representing clothing patterns, such as the
“Mickey Mouse” and “Minions”, have no close semantic connection with the “patterns” on Chinese general corpus. On the other hand, some of these words only appear in the training set, or only appear in the test set.

Table 1. Experimental results of fine-tuning BERT.

| Aspect         | Precision | Recall | F1-score |
|----------------|-----------|--------|----------|
| Fabric         | 100%      | 96.67% | 0.9831   |
| Style          | 95.24%    | 95.24% | 0.9524   |
| Size           | 94.59%    | 97.22% | 0.9589   |
| Color and pattern | 87.80%    | 94.74% | 0.9114   |
| Price          | 100%      | 96.97% | 0.9846   |
| Others         | 100%      | 100%   | 1        |

3.3. Comparative Experiments

To verify the effectiveness of our fine-tuned BERT model, two comparative experiments are conducted. The first comparative experiment is designed for evaluating the effects of fine-tuning. Using the same hyper-parameters, the same pre-training BERT model is trained on the training set to only tune the parameters of the softmax layer, so that the output of pre-training BERT corresponds to the six aspects of clothing reviews. The word2vec+LSTM model is used as the other comparative experiment. The word2vec model is trained on a 268G Chinese corpus including Baidu Encyclopedia text, Sohu News text, and Chinese novels text. Then word2vec is used to generate word vectors for each word in review sentences and word vectors are fed into the LSTM model. The following hyper-parameters are set for the training process: the word vector length is 128, the number of iteration epochs is 30. For the test set, the accuracy, average precision, average recall, and average F1-score of the three different experiments are shown in Table 2. The maximum value of each column is marked in bold. Comparing with word2vec+LSTM, fine-tuning BERT has an improvement of more than 10% for different indicators.

Table 2. The result of three experiments.

| Model                                  | Accuracy | Average precision | Average recall | Average F1-score |
|----------------------------------------|----------|-------------------|----------------|------------------|
| Fine-tuning BERT                        | 96.52%   | 96.70%            | 96.52%         | 0.9657           |
| Pre-training BERT                       | 27.39%   | 25.84%            | 27.39%         | 0.2217           |
| Word2vec+LSTM                           | 85%      | 86%               | 85%            | 0.85             |

4. Conclusion and Extension

In this paper, a fine-tuning BERT based method is proposed to extract implicit aspects. Using labeled data, the hidden layer parameters of pre-training BERT are fine-tuned in a supervised manner to generate fine-tuning BERT. To verify the effectiveness of fine-tuning BERT, some implicit aspect extraction experiments of clothing reviews are conducted. Experimental results show that comparing with the pre-training BERT model and the word2vec+LSTM model, fine-tuning BERT can improve the effect of identifying implicit aspects. To the best of our knowledge, this study is the first to extract implicit aspects with fine-tuning BERT. In the future, more clothing reviews can be used to verify fine-tuning BERT. Moreover, the model can be improved to complete the aspect extraction task where a sentence contains multiple implicit aspects.

Acknowledgements

This research is supported by Natural Science Foundation of Shanghai (No: 18ZR1413200), Institute of Free Trade Zone of SUFE and Shanghai Education Committee (No: 2019110183).
References

[1] Kang, D., & Park, Y. (2014). Based measurement of customer satisfaction in mobile service: Sentiment analysis and VIKOR approach. Expert Systems with Applications, 41(4), 1041-1050.

[2] Xu, H., Zhang, F., & Wang, W. (2015). Implicit feature identification in Chinese reviews using explicit topic mining model. Knowledge-Based Systems, 76, 166-175.

[3] Karagoz, P., Kama, B., Ozturk, M., Toroslu, I. H., & Canturk, D. (2019). A framework for aspect based sentiment analysis on turkish informal texts. Journal of Intelligent Information Systems, 53(3), 431-451.

[4] Sun, L., Li, S., Li, J., & Lv, J. (2014). A novel context-based implicit feature extracting method. In 2014 International Conference on Data Science and Advanced Analytics (DSAA) (pp. 420-424). IEEE.

[5] Hai, Z., Chang, K., Cong, G., & Yang, C. C. (2015). An association-based unified framework for mining features and opinion words. ACM Transactions on Intelligent Systems and Technology (TIST), 6(2), 1-21.

[6] Mohammed, B. (2016, October). Hybrid approach to extract adjectives for implicit aspect identification in opinion mining. In 2016 11th International Conference on Intelligent Systems: Theories and Applications (SITA) (pp. 1-5). IEEE.

[7] Ray, P., & Chakrabarti, A. (2020). A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. Applied Computing and Informatics.

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[9] Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune BERT for text classification?. In China National Conference on Chinese Computational Linguistics (pp. 194-206). Springer, Cham.

[10] Zhu J., Gao G., & F. (2018). Evaluation index system and comprehensive evaluation model of apparel e-commerce sellers. Journal of Textile Research, 39(07), 159-164.