

# Attribute-level Sentiment Analysis Incorporating Auxiliary Information

Huiyang Yin<sup>1</sup>, Jin Hou<sup>1,2,\*</sup>

<sup>1</sup> School of Automation and Information Engineering, Sichuan University of Science & Engineering, Zigong, China

<sup>2</sup> Artificial Intelligence Key Laboratory of Sichuan Province, Yibin, China

\*Corresponding author: Jin Hou (Email: houjin828@sina.com)

**Abstract:** In order to obtain the sentiment tendency of multiple attributes contained in the data reviews, this paper uses an attribute-level sentiment analysis model RBT-BiAtt-GCN, which incorporates auxiliary information and attention mechanism, to analyze the attribute-level sentiment analysis of e-commerce reviews of citrus. The model incorporates auxiliary information, attention mechanism, and GCN network on the basis of the previous one, and the final macro F1 value obtained can reach 78.21%, which achieves good results in comparison with other models.

**Keywords:** Emotional classification, Aspect-level emotional analysis, Attention mechanism.

## 1. Introduction

Attribute-level sentiment analysis is a fine-grained sentiment analysis method whose goal is to identify the sentiment polarity of each attribute in a statement. For example, for the sentence "The property is far away from the city, but the transportation is quite convenient", when the attribute word is "distance", the sentiment polarity is negative; while when the attribute word is "transportation", the sentiment polarity is positive. When the attribute word is "transportation", the sentiment polarity is negative; while when the attribute word is "transportation", the sentiment polarity is positive.

FU [1] et al. used KL scatter to explore the association between a given passage and a topic model. Nadeem Akhtar [2] et al. used a topic model called Mallet. Huaishao Luo [3] et al. proposed a bidirectional dependency tree network that acquires top-down and bottom-up representations and merges them, and then learns feature representations through neural networks for attribute extraction. feature representation for attribute extraction. Yuliya [4] used CRF model for attribute extraction and achieved good results on restaurant and automobile domain datasets. Hu Xu [5] used a bidirectional convolutional neural network for attribute word extraction and achieved significant performance. Peng Chen [6] proposed recurrent neural network based on multiple attention mechanism and applied it to attribute-level sentiment analysis.

In recent years, with the development of Graph Neural Networks (GNN), more attempts have been made to utilize graph neural networks to solve problems in attribute-level sentiment analysis [7].

Ruifan Li [8] and others proposed a dual graph convolutional network model for attribute sentiment classification task, capturing syntactic knowledge and

semantic relevance respectively, and proposed two regularization methods to constrain the model, which led to a better generalization ability. Zeguan Xiao [9] proposed a BERT-based graph convolutional neural network for attribute-level sentiment analysis. Ruyan Wang et al.

proposed a multi-interaction graph convolutional network model, which takes the positional features of each word in the text as the input to each layer of the GCN, and at the same time weights the adjacency matrices of the GCN using the dependent syntactic tree, and the syntactic distance features of the text words, and handles the semantic and grammatical information between words through the interactions of semantic and syntactic rules.

## 2. Related Theory and Technology

### 2.1. BERT

Pre-training models is a current hot topic in NLP, where unsupervised learning equips models with certain prior knowledge to assist downstream models in semantic representation of the target task, thus avoiding overfitting.

BERT learns highly enriched language representations by performing pre-training tasks of Masked Language Model (MLM) and Next Sentence Prediction (NSP) on a large-scale text corpus. These pre-trained language representations can be fine-tuned and applied to various types of natural language processing tasks, such as sentiment analysis, named entity recognition, etc., where they show excellent performance.

BERT is known for its strong representational capabilities and versatility, and can be adapted to a wide range of natural language processing tasks and achieve excellent performance in many challenging tasks. In addition, with its open source and easy-to-use features, BERT has become one of the important milestones in the field of natural language processing, and the basic structure of BERT is shown in Figure 1.

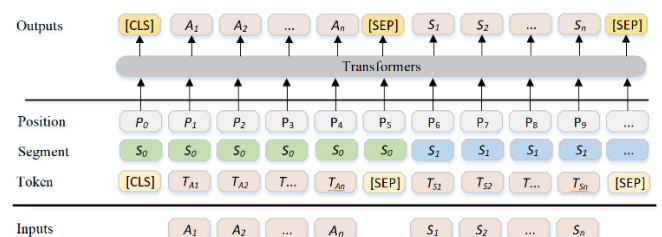


Figure 1. Structure of BERT

## 2.2. ROBERTA

ROBERTA is an improved pre-trained language model introduced by Facebook AI in 2019 to optimize and enhance the original BERT model. ROBERTA is based on the Transformer architecture and uses large-scale unsupervised training to learn language representations. Unlike BERT, ROBERTA uses larger batch training, longer training time, and more data for pre-training to achieve better results in various natural language processing tasks. In addition, ROBERTA employs dynamic masking technique to re-select the mask position randomly in each iteration instead of fixing the mask position like BERT, and experimentally demonstrates that the NSP task has no effect as selection removal in natural language processing.

## 2.3. BIGRU

In 2014 K Cho et al. proposed gated recurrent unit (GRU) based on LSTM. The gated recurrent unit is a recurrent neural network structure similar to a long short-term memory network, which is designed to solve the gradient vanishing and information transfer problems in modeling long sequence data.

The GRU consists of two important gating units: the Reset Gate and the Update Gate. The Reset Gate controls how to integrate past information and the Update Gate controls how to integrate new input information. GRU also uses "candidate values" to update the current internal state while preserving the old state information.

Compared with LSTM, GRU has a simpler structure and higher computational efficiency, and therefore has better performance in some scenarios. GRU is widely used in natural language processing, speech recognition, machine translation and other tasks, and has become one of the most important models for processing sequential data. Its design concept is to optimize the information flow through the gating mechanism and improve the modeling ability for long sequence data. The internal structure of the GRU is shown in Figure 2.

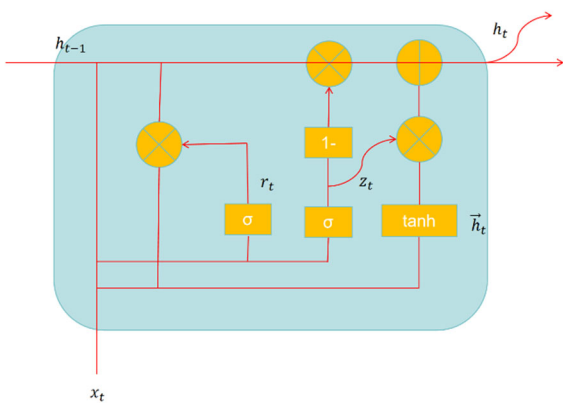


Figure 2. Internal framework of GRU

Here,  $r$  and  $z$  control the reset gate and update gate, respectively; their states are jointly determined by the hidden state of the previous node and the inputs of the current node.  $\sigma$  represents the Sigmoid function, which is used to convert the data to a value between 0 and 1, and serves as a gating signal.  $r$  and  $z$  are expressed as in Eqs.1 and 2:

$$r = \sigma(w_r[x_t, h_{t-1}]) \quad (1)$$

$$z = \sigma(w_z[x_t, h_{t-1}]) \quad (2)$$

The reset gate determines how much of the historical information will be forgotten by the time it passes to the next time step and thus is used to reset the stored information. The GRU directly utilizes the hidden unit to record the historical state and adjusts the trade-off ratio between the current information and the memorized information by using the reset gate in order to generate the new stored information and continue to pass forward. The output of the reset gate ranges between  $[0, 1]$ , when the output is 0 it means that all the memorized information is reset, and when the output is 1 it means that all the memorized information is passed all the way through. After obtaining the gating information,  $r$  is first used to control the generated "reset" data, which is processed in the following equations 3 and 4:

$$h'_{t-1} = h_{t-1} \odot r \quad (3)$$

$$h' = \tanh(w[x_t, h_{t-1}]) \quad (4)$$

The update gate determines how much past information will be passed on to the future, similar to the input and forget gates in LSTM, and it is used to compute the output of the hidden state at the current moment. The data processing of the update gate is similar to that of the reset gate, but the values and roles of the weight matrix are different. By utilizing  $z$  for updating, the update expression is shown in Eq.5:

$$h_t = (1 - z) \odot h'_{t-1} + z \odot h' \quad (5)$$

Unidirectional GRU can only capture forward sequence features and cannot acquire backward information of text. Therefore, some scholars have proposed BiGRU to capture long text sequence features. BiGRU is a neural network structure that synthesizes both forward and backward information, combining the advantages of bi-directionality and gating mechanism, which helps to process sequence data efficiently and capture the long term dependencies in it. The structure of BiGRU is shown in Figure 3.

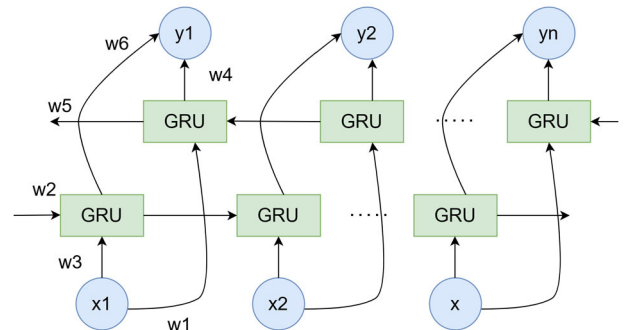


Figure 3. Framework of BiGRU

## 2.4. TEXTCNN

The input to TextCNN is usually converting the text into a sequence of word embedding vectors and then feeding these vectors into a convolutional layer for feature extraction. Convolutional operations help in capturing combinations of words in different length ranges, thus learning local features



Context-Position Attention: on the basis of the new attribute word representation, combined with the position encoding information, after modeling through the attention mechanism, an attribute representation that integrates lexical, positional, and contextual information is finally obtained, which is used for subsequent input to graph convolutional neural network for sentiment classification.

GCN layer:

In order to capture the sentiment dependencies between attributes in the comment text, each attribute in the comment text is considered as a node, and then edges are constructed based on the adjacencies between the attributes, and these edges represent the sentiment dependencies between the attributes. The generated graph structure includes these nodes and edges as inputs to the graph convolutional neural network (GCN).

### 3.2. Criteria for evaluation

This chapter focuses on measuring model performance using F1 (F1-Score) values as well as macro-averaged F1 (Macro\_F1) values. the F1 values are calculated using the following formula:

$$F1 = \frac{2 * P * R}{P + R} * 100\% \quad (10)$$

where P precision rate and R is the recall rate.

$$P = \frac{TP}{TP + FP} * 100\% \quad (11)$$

$$R = \frac{TP}{TP + FN} * 100\% \quad (11)$$

Where TP means that the prediction is positive and the actual value is also positive;FP means that the prediction is positive but the actual value is negative; FN means that the prediction is negative but the actual value is positive.

The Macro\_F1 calculation formula is shown in 12.

$$Macro\_F1 = \frac{1}{X} \sum_{i=1}^X F1_i \quad (12)$$

### 3.3. Data set

This chapter first experiments with the model using the publicly available dataset AI Challenger2018 to validate the effectiveness of the model. This dataset is user reviews in the field of food and beverage, which is often used in Chinese attribute-level sentiment analysis tasks, and can be used very reliably in related research. The dataset is divided and has been labeled with 4 types of sentiment tendencies for 20 fine-grained attributes, with the sentiment tendency labels: {Unmentioned: -2, Negative: -1, Neutral: 0, Positive: 1}. In the experiments in this chapter, four of these fine-grained attributes have been selected for the study.

### 3.4. Baseline modeling experiments

ATAE-LSTM: a neural network model for attribute-level sentiment analysis, by introducing an attentional mechanism, the model is able to dynamically adjust the level of attention paid to the target and related words in order to capture the

emotional connection between the target and the expression more efficiently. The LSTM layer is used to process sequential data, which helps the model to understand long-distance dependencies.

IAN: interactive attention network for ATSA tasks, based on LSTM and attention mechanisms.

AEN: the model mainly consists of an embedding layer, an attentional coding layer, an attention layer for specific aspect words and an output layer.

BERT-LSTM: By combining BERT and LSTM to increase the utilization of textual information, and finally the output results of the fully connected layer are used to judge the sentiment classification. Table 1 shows the experimental results of each model. In Table 1 P, U, and N are negative, neutral, and positive, respectively.

Table 1. Experimental results for each model

model	P-F1 (%)	U-F1 (%)	N-F1 (%)	M-F (%)
ATAE-LSTM	76.83	45.51	71.81	64.72
IAN	70.34	55.14	61.49	62.32
AEN	73.96	45.32	70.79	63.36
BERT-LSTM	84.10	54.11	81.87	73.36
RBT-BiAtt-GCN	89.26	59.15	86.22	78.21

As can be seen from Table 1, the BERT-LSTM and RBT-BiAtt-GCN results are greatly improved compared to the other three models. This is obviously the pre-training model plays a role, which shows that the use of linguistic pre-training model can really improve the performance of the model in sentiment analysis. The RBT-BiAtt-GCN, on the other hand, is optimal in all metrics, indicating that the model used in this chapter works better in the attribute-level sentiment analysis experiments.

### 3.5. ablation experiment

In order to better study the effectiveness of the RBT-BiAtt-GCN (RBG) model, this paper designs relevant ablation experiments to remove the lexical information, positional coding information, and GCN layer in the model while keeping the rest of the model unchanged, respectively. They are denoted as RBG/Pos-Tag, RBG/Postion, and RBG/GCN, respectively. The experimental results are shown in Table 2.

Table 2. Model ablation experiments

Model	Marco acc (%)	Marco F1 (%)
RBG/Pos-Tag	80.07	74.10
RBG/Postion	79.08	73.24
RBG/GCN	77.62	72.51
RBG	80.23	78.21

As can be seen from Table 2, the results of all three sets of ablation experiments have decreased compared to the final model. After removing the lexical labeling information, the metrics of the model have decreased, indicating that the lexical labeling information can play a role in attribute-level sentiment analysis, and the metrics have decreased more after removing the positional coding information, indicating that the positional coding information plays a greater role compared to the lexical labeling information. After removing the GCN, the model is the least effective, indicating that the GCN has a pivotal position among the three modules.

### 3.6. GCN Layer Comparison Experiment

Based on the results presented in Figures 6, experiments on the dataset show that the model performs best when the number of layers of the GCN is 2. With a gradual increase in the number of layers, the number of model training parameters increases leading to overfitting and thus a decrease in performance on the dataset.

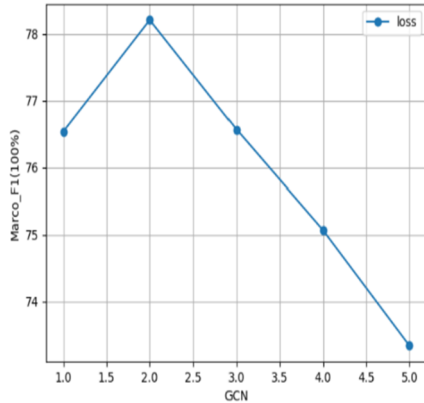


Figure 5. GCN layer comparison

## 4. Conclusion

In this paper, a RBT-BiAtt-GCN attribute-level sentiment analysis research model with fusion of auxiliary information is proposed, in order to ensure the reliability of the experiments, firstly, the experiments are carried out by using the public dataset, the model is compared with other models, and then the ablation experiments of the model are designed, which proves that the model can achieve better results in attribute-level sentiment analysis research, and proves that the GCN and the auxiliary information have the effect of improving the attribute-level sentiment analysis.

## References

- [1] FU X, LIU G, GUO Y, et al. Multi-aspect blog sentiment analysis based on LDA topic model and hownet lexicon; proceedings of the International Conference on Web Information Systems and Mining, F, 2011 [C]. Springer.
- [2] AKHTAR N, ZUBAIR N, KUMAR A, et al. Aspect based sentiment oriented summarization of hotel reviews [J]. Procedia computer science, 2017, 115: 563-71.
- [3] LUO H, LI T, LIU B, et al. Improving aspect term extraction with bidirectional dependency tree representation [J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2019, 27(7): 1201-12.
- [4] RUBTSOVA Y, KOSHELNIKOV S. Aspect extraction from reviews using conditional random fields; proceedings of the International Conference on Knowledge Engineering and the Semantic Web, F, 2015 [C]. Springer.
- [5] XU H, LIU B, SHU L, et al. Double embeddings and cnn-based sequence labeling for aspect extraction [J]. ar Xiv preprint arXiv:180504601, 2018.
- [6] CHEN P, SUN Z, BING L, et al. Recurrent attention network on memory for aspect sentiment analysis; proceedings of the Proceedings of the 2017 conference on empirical methods in natural language processing, F, 2017 [C].
- [7] PHAN H T, NGUYEN N T, HWANG D. Convolutional attention neural network over graph structures for improving the performance of aspect-level sentiment analysis [J]. Information Sciences, 2022.
- [8] LI R, CHEN H, FENG F, et al. Dual graph convolutional networks for aspect-based sentiment analysis; proceedings of the Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), F, 2021 [C].
- [9] XIAO Z, WU J, CHEN Q, et al. BERT4GCN: Using BERT Intermediate Layers to Augment GCN for Aspect-based