International Journal of Computer Applications Technology and Research
Volume 12–Issue 06, 08-11, 2023, ISSN:-2319–8656
DOI:10.7753/IJCATR1206.1002

An Aspect-Level Sentiment Analysis Approach Based on BERT and Attention Mechanism

Jinbo Liang
ChengDu University Information of Technology
ChengDu, China

Hao Peng
ChengDu University Information of Technology
ChengDu, China

Yuhao Zhan
ChengDu University Information of Technology
ChengDu, China

Abstract: In recent years, with the rapid increase in the number of comment texts on social media, more and more researchers have been studying the emotional tendencies in texts. In traditional sentiment analysis methods, there are still problems such as weak semantic dependency relationships and loss of semantic information caused by one-way networks. The paper proposes a sentiment classification method based on BERT and attention mechanism, which uses BERT as the word embedding model to obtain word vectors containing more semantic information, there by mitigating the impact of semantic sparsity , the feature extraction layer uses bidirectional gated units to extract hidden vector information. The bidirectional network avoids the loss of forward semantics. The semantic interaction layer models the aspect words and context at the same time, and enhances the semantic dependency relationship between texts through interactive attention based on constructing global semantics. The experimental results show good performance.

Keywords: BERT; sentiment classification; attention mechanism; global semantics; semantic interactive

1. INTRODUCTION

In recent years, deep learning has been widely applied in the field of natural language processing. Bengio et al.[1]were the first to introduce neural networks into language models. Zhang and his colleagues et al.[2] transformed the input of convolutional neural networks into high-dimensional text data, and introduced the bag-of-words model into the convolutional layer of the neural network to obtain the Bow-CNN and Seq-CNN models. Kim et al. [3] proposed the Test -CNN text classification model, which trained the model using different types of word embeddings. Nguyen et al. [4] conducted sentiment analysis on comment text using RNNs. Liu et al. [5] proposed three information sharing mechanisms based on RNNs, respectively modeling specific text classification tasks and the shared layers. Wang et al. [6] argued that the sentiment polarity of a sentence should not be determined solely by its content , but is highly related to aspect words in the sentence. Akhter et al.[7] proposed a new deep learning architecture that uses CNNs to obtain sentiment word embeddings. Bahdanau et al.[8] introduced an attention mechanism in text translation. Peng et al.[9] used a multi-level attention mechanism to capture sentiment features that are far away in the text, while paying attention to semantic information of multiple aspect words. Ma [10] employed an attention mechanism to process textual information, which can automatically learn the importance of different parts of the text and conduct sentiment analysis based on their importance.

2. RELATED WORK

In this paper, we propose a BERT-based aspect-level sentiment analysis model with attention mechanism, as shown in Figure 1.

The model architecture consists of a word embedding layer, feature extraction layer, semantic interaction attention layer, and output layer.

2.1 Word Embedding

The word embedding layer utilizes the pre-trained BERT model to transform the input context text $S^t = \{S_1^t, S_2^t, \ldots, S_m^t\}$and aspect word $S^t = \{S_1^t, S_2^t, \ldots, S_m^t\}$into word vectors $W^c = \{W_1^c, W_2^c, \ldots, W_n^c\}$ and $W^t = \{W_1^t, W_2^t, \ldots, W_n^t\}$.

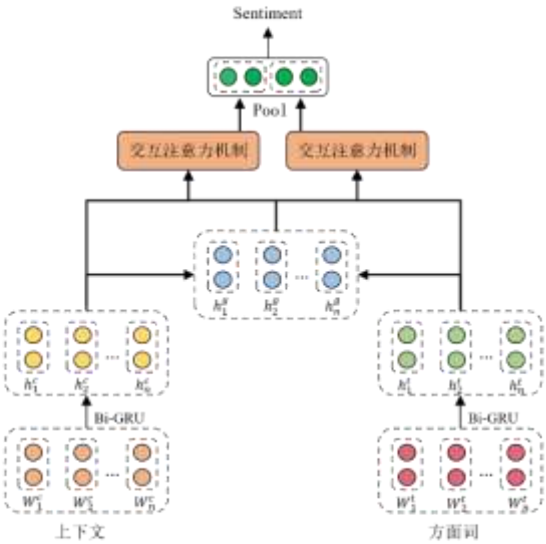


Figure.1 The semantic interaction model based on BERT and attention mechanism.

2.2 Feature Extraction

After obtaining the word vectors for the context text and aspect words, it is necessary to extract the implicit feature information in the word vectors. The word vectors $W = \{W_1, W_2, \ldots W_n\}$generated by the pre-trained BERT model [11] can be used as the input of the Bi-GRU network. As shown in the figure, the GRU [12] network encodes the input vectors in two directions, forward and backward, to obtain the forward hidden state encoding $\overrightarrow{h} = \{\overrightarrow{h_1}, \overrightarrow{h_2}, \ldots \overrightarrow{h_n}\}$ and the backward hidden state encoding $\overleftarrow{h} = \{\overleftarrow{h_1}, \overleftarrow{h_2}, \ldots, \overleftarrow{h_n}\}$.Finally, the two directions' sequences are concatenated to obtain the final representation of the hidden state $h = \{h_1, h_2, \dots h_n\}$, which is calculated as follows:

$$\vec{h}_i = \overrightarrow{\mathrm{GRU}}(x_i) \qquad (3\text{-}1)$$

$$\overleftarrow{h}_i = \overleftarrow{GRU}(x_i) \qquad (3\text{-}2)$$

$$h_i = \{\vec{h}_i ; \overleftarrow{h}_i\} \qquad (3\text{-}3)$$

2.3 Feature Extraction

The interactive attention module consists of three parts. The semantic vectors of the context and aspect words are used as the Key part of the attention mechanism, and the concatenated global semantic vector is used as the Query part of the attention mechanism. The relevance between Q and K is computed to represent which part of the context and aspect words is more important to the global semantic representation. Then, the attention values of each part are obtained by multiplying and summing them with the corresponding Value in the attention mechanism. Finally, they are concatenated, linearly transformed, and fed into the classification function to obtain the final sentiment polarity.

Firstly, the vector matrices of the context and aspect words are concatenated and averaged separately to obtain the global semantic vector and the initial representations of both.

$$e^g = \{e^t ; e^c\} \qquad (3\text{-}4)$$

$$v^c = \frac{e_i^c}{n} \qquad (3\text{-}5)$$

$$v^t = \frac{e_i^t}{m} \qquad (3\text{-}6)$$

Secondly, the attention mechanism is used to learn the attention between global semantic vector and context/aspect words. The initial representations v^t , v^c of the vector, and the hidden state vector e^g are taken as inputs to obtain the probability matrix of e^g for v^t and v^c which is the attention distribution. The calculation formula is as follows:

$$\alpha_i \frac{\exp(f_s(e_i^g, v^c))}{\sum_{j=1}^{n} \exp(f_s(e_j^g, v^c))} \qquad (3\text{-}7)$$

$$\beta_i \frac{\exp(f_s(e_i^g, v^t))}{\sum_{j=1}^{m} \exp(f_s(e_j^g, v^t))} \qquad (3\text{-}8)$$

In the formula, $f_s()$ is an addition function in attention mechanism, which calculates the correlation between Query and Key. There are multiple forms of addition functions in attention mechanism, and we choose the dot-product model here, which is expressed as:

$$f_s(k_i, q_i) = \tanh([k_i ; q_i], W_s) \qquad (3\text{-}9)$$

After obtaining the weight coefficients of $Value_i$ namely α_i and β_i , the attention representations of the two can be obtained by weighted summation:

$$a_l = \sum_{i=1}^{n} a_i e_i^g \qquad (3\text{-}10)$$

$$b_l = \sum_{i=1}^{m} \beta_i e_i^g \qquad (3\text{-}11)$$

Concatenating the attention obtained by the interaction between global text and context and aspect words, the resulting attention representation is:

$$h_l = \{a_l ; b_l\} \qquad (3\text{-}12)$$

The concatenated matrix vector is projected to the target matrix space of class C through a fully connected layer, and the final representation of the global semantic interaction layer is obtained:

$$o = W_o^T h_l + b_h \qquad (3\text{-}13)$$

3. EXPERIMENTS

3.1 Setup

In the experiment, publicly available datasets SemEval2014 Task 4 and Twitter were used for aspect-level sentiment analysis task. The number of data samples in the training set and test set are shown in Table 3-2.

Table3-2 Data sample statistics

Datasets	Positive		Neural		Negative	
	Train	Test	Train	Test	Train	Test
Restaurant	2164	728	637	196	807	196
Laptop	994	341	464	169	870	128
Twitter	1561	173	3127	346	1560	173

The evaluation metrics for the experimental results were accuracy and F1 score. Accuracy refers to the proportion of correctly classified samples in all classification samples of the sentiment analysis model. F1 is the harmonic mean of precision and recall, which is mainly used to evaluate the overall performance of the sentiment analysis model. The

Table 3-6: Experimental results comparison of various models on the datasets.

Model	Restaurant		Laptop		Twitter	
	Acc (%)	F1(%)	Acc (%)	F1(%)	Acc (%)	F1(%)
LSTM	74.30	-	66.50	-	66.50	64.72
TD-LSTM	75.60	-	68.13	-	70.80	69.00
ATAE-LSTM	77.20	-	68.70	-	-	-
IAN	78.60	-	72.10	-	-	-
MGAN	81.25	71.94	75.39	72.47	72.54	70.81
BERT-BASE	82.66	74.13	79.04	73.02	73.02	71.43
Clove-GSIA	81.52	70.89	78.87	72.93	73.41	70.79
BERT-GSIA	**83.04**	**74.83**	**80.03**	**75.79**	**74.28**	**72.42**

The accuracy and F1 values of the benchmark model in the table are taken from the original text, where the "-" symbol indicates that the relevant parameter was not mentioned in the original article.

higher the F1 score, the better the performance of the model. The calculation formula is as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3\text{-}16)$$

$$F1 = \frac{2 \times pression \times recall}{pression + recall} \qquad (3\text{-}19)$$

3.2 Experimental Result and Analysis

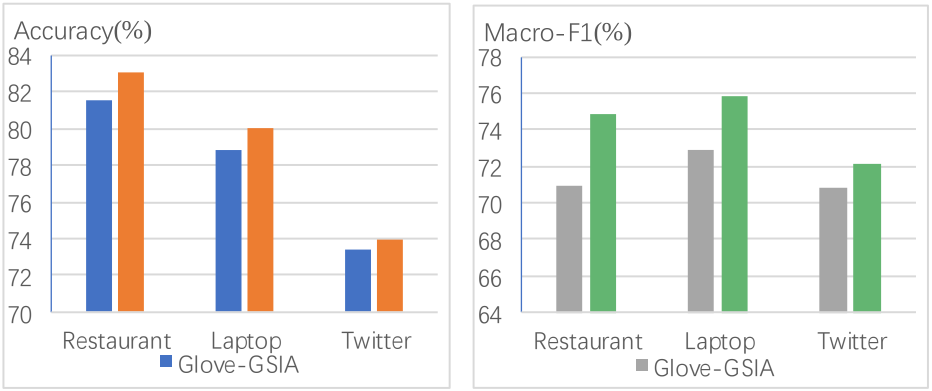
This experiment was conducted under the PyTorch deep learning framework. The model learning rate was set to 2e-5, the regularization coefficient was set to 1e-5, the maximum sentence length was 85, and the batch size for input samples was 16.

The experimental results of the global semantic interaction algorithm model based on BERT and attention mechanism proposed in this chapter are compared with those of related models as follows3-6:

Based on the Glove word embedding model, it performs well. The reason for this is that the bidirectional gated units are used in the feature extraction layer to obtain the forward semantic information lost in the unidirectional network, and introducing the fusion of aspect word and context semantic features into the interactive attention mechanism can enhance the correlation between aspect words and texts, enabling the aspect words and target sentiment words to capture each other's dependency information and gain more weight. In terms of accuracy improvement, there is a significant improvement on the Laptop dataset because it contains more implicit expression sentiment samples, while the Restaurant dataset has more surface-level emotions, proving that the proposed model can better mine deep-seated information.

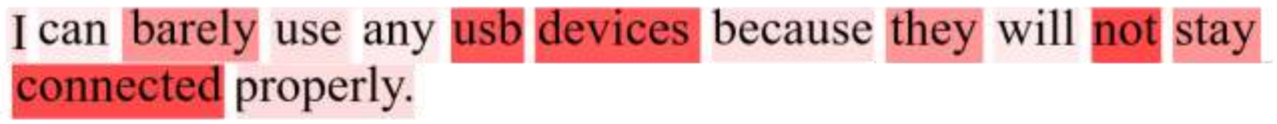
The BERT-BASE benchmark model performs better than the Glove-GSIA on the Restaurant and Laptop datasets, but has poorer accuracy on the Twitter dataset. The reason for this is that the multi-layer Transformer mechanism in BRET can capture more semantic information to convert into word vectors and perform classification. On the other hand, BERT-GSIA uses BERT as a word embedding model, and its performance on all three datasets is better than the BERT-BASE model, proving that the model can effectively utilize the bidirectional network to obtain semantic information and enhance the correlation between context and aspect words through semantic interaction attention mechanism in downstream tasks such as sentiment analysis. The comparison of different word embedding models is shown in the following figure2:

3.3 Attention Visualization



(a) (b)

Figure 3-9 The impact of different word embedding models on the model.



I can barely use any usb devices because they will not stay connected properly.

Figure 3-11 Attention Weight Visualization

From the figure, it can be seen that when analyzing the sentiment of the aspect term, more weight coefficients are assigned to the corresponding target sentiment words, which also have a greater impact on the polarity judgment of the sentiment. The colors of "not" and "connected" in the text are darker than other words, indicating that the attention mechanism pays more attention to the target sentiment words of the aspect term and assigns more weight. However, the polarity of "connected" is ultimately negated by "not", which enables the model to reasonably judge the sentiment orientation of "usb devices". This also demonstrates the effectiveness of the attention mechanism in the model.

4. CONCLUSIONS

In this paper, we propose a new method for aspect-based sentiment analysis. By obtaining more word vector information, we model the aspect term and context separately, and construct global semantic information to achieve interaction between the information, enhancing the connection between semantics. Through comparison, our proposed method outperforms relevant baseline models.

5. REFERENCES

[1] Y Bengio*, Ducharme R , Vincent P . A Neural Probabilistic Language Model[J]. 2001.

[2] Johnson R , Tong Z . Effective Use of Word Order for Text Categorization with Convolutional Neural Networks[J]. Eprint Arxiv, 2014.

[3] Kim Y . Convolutional Neural Networks for Sentence Classification[J]. Eprint Arxiv, 2014.

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

[5] Liu P , Qiu X , Huang X . Recurrent Neural Network for Text Classification with Multi-Task Learning: AAAI Press, 10.48550/arXiv.1605.05101[P]. 2016.

[6] Wang Y , Huang M , Zhu X , et al. Attention-based LSTM for Aspect-level Sentiment Classification[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. 2016.

[7] Akhtar S , Kumar A , Ekbal A , et al. A Hybrid Deep Learning Architecture for Sentiment Analysis[C]// COLING. 2016.

[8] Bahdanau D , Cho K , Bengio Y . Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.

[9] Chen P , Sun Z , Bing L , et al. Recurrent Attention Network on Memory for Aspect Sentiment Analysis[C]// Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017.

[10] Ma D , Li S , Zhang X , et al. Interactive Attention Networks for Aspect-Level Sentiment Classification[J]. 2017.

[11] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding[J]. 2018.

[12] Cho K, Merrienboer B V, Gulcehre C ,et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.