Fusion Local and Global Aspect-based Sentiment Analysis

Yusheng Zhou, Lei Li and Jie Li

April 20, 2023

Fusion Local and Global Aspect-based Sentiment Analysis

Abstract—**Aspect-based sentiment classification is an important task in natural language processing research, and in response to the fact that most studies at this stage ignore the influence of contextual semantic information on the sentiment polarity of aspect words, the BiLSTM-LCF model proposed in this paper combines local aspect word feature extraction and global contextual semantic information extraction based on Bi-directional Long Short-Term Memory (BiLSTM), and after a multi-headed attention mechanism to enhance the local aspect word sentiment representation. Comparative experiments were conducted on the restaurant and laptop datasets of the SEMEVAL2014 evaluation task. The experimental results show that the model proposed in this paper achieves good classification results in the aspect-level sentiment analysis task of text reviews. The method provides a new idea for ABSA task development.**

Keywords—sentiment analysis,BiLSTM,interactive attention, BERT,aspect-level

I. Introduction

With the booming development of Internet technology and the corresponding rapid development of e-commerce platforms such as Amazon and Taobao, many users will evaluate the after-sales service of the store, the quality of the items and other factors after purchasing the goods, and these evaluations will often affect the desire of other users to buy. Merchants can use natural language processing technology to analyze the sentiment of users' online reviews to help them understand what needs to be improved, thus improving the shopping experience for users.

Aspect-based sentiment analysis (ABSA) [1] is a fine-grained task compared to traditional sentiment analysis, where user reviews of a product generally need to consider multiple aspects, e.g., given a review of an item: "The design of this dress is beautiful, but the shipping speed is too slow." Users express both positive and negative sentiments about the "dress" and "shipping speed". Merchants can then improve their shipping speed and other factors based on the analysis of the review text.

In the traditional research on text sentiment classification tasks, most of the methods based on sentiment dictionaries and machine learning are used; for example, Reb et al. incorporated sentiment analysis into a method based on SVM combined with sliding windows to improve the accuracy of prediction. However, the disadvantage of the machine learning approach is that the sentiment analysis of text often does not fully utilize contextual information, which has an impact on the accuracy of classification results. With the continuous development of deep learning, more and more researchers have started to study natural language processing problems using neural network algorithms, and a lot of useful work has been done on sentiment analysis problems.Shinhyeok [2] proposed a long and short term memory (LSTM) model combining attention mechanisms, which utilizes an attention mechanism of embedding aspects to focus on different parts of a sentence separately. Zeng et al [3] proposed a Local Context Focus (LCF) model, which uses word distances in sentence sequences as Semantic-Relative Distance (SRD) to obtain local contextual representations. The model notes that the sentiment polarity of an aspect is more relevant to context words close to itself, and that context words far from the aspect words may negatively affect the prediction accuracy of the polarity of a particular aspect.Wu[4] applied the capsule network to the sentiment analysis task by referring to the capsule network in the image processing field, and effectively reduced the problem of information loss in the model during feature iteration. Recently pre-trained models based on Bert, PLM, XL-NET, ALBERT, etc. have made a splash in several tasks in NLP, by making full use of the large-scale monolingual corpus compared to traditional methods using language model pre-training, which can model multiple meanings of a word, the process of pre-training using language models can be regarded as a sentence-level contextual word representation, and achieve better sentiment analysis by fine-tuning the pre-trained model.Wu et al [5] used Context-Aware Self-Attention Networks as the feature extraction layer of the model and proposed the CG-BERT model (Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis, CG-BERT), which obtained better results than the basic BERT model, and the introduction of context-aware self-attention networks enabled deep learning models to better extract the global semantic features of the input sequences, with cutting-edge results on several datasets.

Inspired by the above research, this paper proposes a BiLSTM-LCF model that fuses contextual and local features, which obtains rich semantic encoding by pre-training Bert, then captures features that are semantically more relevant to aspect words through local attention layers, uses BiLSTM to capture contextual semantic features, and passes the fused features through a multi-headed attention mechanism to capture the potential interaction between aspect and context, achieving improved accuracy of aspect-level sentiment analysis.

II. Related Works

ABSA can be divided into traditional machine learning methods and the latest deep learning methods, with the application and development of neural networks in the field of natural language processing, the generalizability and scalability of deep learning-based methods have been improved, and the application in sentiment classification tasks has achieved considerable results. The generalizability and scalability of the method have been improved, and the application in sentiment classification tasks has achieved considerable results. Among them, recurrent neural networks and convolutional neural networks are the commonly used deep neural network models.

Tang D Y et al. [6] used the Long Short-Term Memory (LSTM) model to capture the characteristics of the relationship between aspect words and context, and divided the text into two parts with aspect words as the center. Two LSTMS were used respectively to encode the two parts. The target dependent LSTM and target connection LSTM models are proposed. The aspect words and context words are input into the LSTM for classification. Due to the linear structure of RNN, LSTM and gated cycle unit tend to ignore the remote input, and it is difficult to capture the context representation that is far away from aspect words. In order to solve this problem, Wang splices the hidden layer output of LSTM with aspect words, and through the attention mechanism, The model can extract the context related to aspect words from complex texts and improve the accuracy of fine-grained emotion classification. Chen Hong et al. [7] used bidirectional LSTM to model context and aspect words respectively, and used interactive attention module to capture the relationship between the two. Aiming at overlapping feature expressions in aspect level emotion classification tasks, Xu Zhidong et al. [8] used sequence convolution to extract context and aspect word features respectively, and proposed an aspect level emotion classification model based on capsule network. Song Ting et al. [9] the text is divided into multiple aspects words contain statements, using regional convolution sampling the local characteristics of different clause, puts forward using long since attention (multi - head self attention, MHSA) mechanisms respectively learning context the relationship between the internal characteristics and context and word.

Zeng et al. believed that the emotional polarity of an aspect word is more closely related to the Local Context words near itself, while the context words far away from the aspect words may have a negative impact on the prediction of its emotional polarity, so they proposed the Local Context Focus (LCF) method of the aspect words.Peng[10], Ma, Li and Cambria (2018) proposed the methods for the Chinese language APC task, which conducted the APC task at the aspect level via three granularities. Two fusion methods for the granularities in the Chinese APC task are introduced and applied. Empirical results show that the proposed methods achieved promising performance on the most commonly used ABSA datasets and four Chinese review datasets. Meanwhile, a joint framework aimed to aspect sentiment classification subtask and aspect-opinion pair identification subtask is proposedby Chen and Huang (2019)[11],in which the external knowledge are considered and put into the network to alleviate the problem of insufficient train data. The gated alternate neural network (GANN) Liu [12]and Shen (2019) proposed for APC task aimed to solve the shortcomings of traditional RNNs and CNNs. The GANN applied the gate truncation RNN (GTR) to learn the aspect-dependent sentiment clue representations. Zeng[13], Ma, Chen and Li (2019b) proposed an end-to-end neural networkmodel for the ABSA task based on joint learning, and the experimental results on a Chinese review show that the proposed model works fine while conducting ATE and APC subtask simultaneously.BERT-SPC is the BERT text pair classification model, it is a variation model of BERT and is adapted to solve the ABSA task in Song et al. (2019) and achieve high performance.. BERT-ADA Rietzler, Stabinger, Opitz and Engl (2019)shows that although the pre-trained model based on a large universal corpus, and is easy to be applied to most tasks and improve performance. Still, it is not task-specific. For specific tasks, if the pre-trained BERT is adapted to specific tasks through the fine-tuning process on a task-related corpus, the task performance can be further improved.

III. MODEL ARCHITECTURE

The aspect word sentiment classification model in this paper is shown in Fig.1, macroscopically divided into three input layers, feature extraction layer and output layer. In this paper, two models are trained separately, aspect word local feature extraction and global above feature extraction, and finally feature fusion is performed to obtain the sentiment polarity of corresponding aspect words.

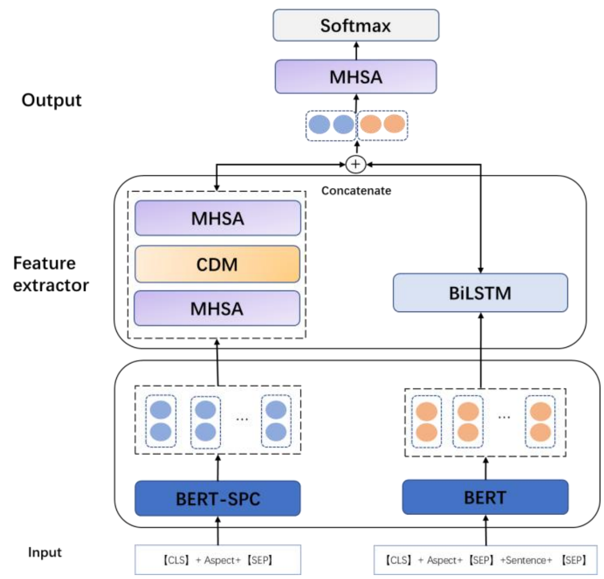

Fig.1.Model structure of BiLSTM-LCF

3.1 Input

In the BiLSTM-LCF model, in order to obtain better encoding features, the BERT-SPC input sequence in this paper is processed as "[CLS]" + context sequence + "[SEP]" + aspect word sequence + "[SEP]",the BERT input sequence is "[CLS]" + context sequence + "[SEP]". The "[CLS]" and "[SEP]" are 2 types of markers used by BERT, where "[CLS]" is the classification marker and "[SEP]" is the classification marker. "[SEP]" is a separator token, which is used to separate different sequences when multiple sequences are entered. The original input sequences are processed into the input format required by BERT for text classification tasks, thus maximizing the effect of BERT.

3.2 Feature Extractor

3.2.1Bi-directional Long Short-Term Memory(BiLSTM)

After pre-training Bert encoding for text sequences, the input hidden layer sequences possess better semantic information. In order to obtain richer contextual semantic information, the long-range semantic information can be better captured by the combination of forward LSTM and reverse LSTM of bi-directional long short-term memory network (BiLSTM).Long Short-Term Memory (LSTM) [14] is based on recurrent neural network by setting forgetting gates, input gates and output gates to selectively The Long Short-Term Memory (LSTM) is based on the recurrent neural network and selectively forgets the meaningless information in the past and retains the new useful information by setting the forgetting gate, input gate and output gate. Compared with recurrent neural networks, LSTM can better capture longer distance dependencies.The network structure of BiLSTM model can be represented by Eqs. (1) to (3). Where, Equation (2) represents the state of the moment forward LSTM layer, Equation (1) represents the state of the moment backward LSTM layer, corresponding to the word embedding vector, corresponding to the weight, and the activation function, the BiLSTM output vector is O^g。.

$$h_{tr} = f(w_1 \cdot x_t + w_2 \cdot h_{t-1r}) \tag{1}$$

$$h_{tl} = f(w_3 \cdot x_t + w_5 \cdot h_{t+1r}) \tag{2}$$

$$O^g = f(w_4 \cdot h_{tl} + w_6 \cdot h_{tl}) \tag{3}$$

3.2.2 Semantic-Relative Distance(SRD)

SRD is generally divided into sequence distance and dependency distance, where sequence distance refers to the absolute distance between two words in a sentence, and dependency distance refers to the shortest distance between two words in the corresponding nodes in the syntactic dependency tree. In this paper, the sequence distance is chosen as the semantic correlation distance, which is experimentally verified to have better prediction effect. Suppose the text sequence is M $\{M = m_1, m_2, m_3, \ldots, m_i, m_{i+1}, \ldots, m_{i+m-1}, m_{i+m}, \ldots, m_n\}$, where $\{m_i, m_{i+1}, \ldots, m_{i+m-1}\}$ are sequences of aspects, for any word the SRD is calculated as follows.

$$SRD_i = \left|j - (i + \frac{m}{2})\right| - \frac{m}{2} \tag{4}$$

3.2.3 Context-features Dynamic Mask(CDM)

By calculating the semantic correlation distance between each word in the sentence and the aspects, and then according to the larger SRD with the aspects indicates that it has less influence on the affective tendency of the aspects, the weight it occupies should be appropriately reduced, and the highest weight should be given to the word with small semantic correlation distance.

The following mask vector M is designed in our research:

$$m_i = \begin{cases} E & SRD_i \le a \\ \dfrac{2a - SRD_i}{n} \cdot E & SRD_i > a \end{cases} \tag{5}$$

$$M = [m_1, m_2, m_3, \ldots, m_n] \tag{6}$$

$$O_{cdw} = E^l \cdot M \tag{7}$$

where is the all-one vector,is the dimension of the context hidden vector,is the threshold of SRD,n is the length of the sentence,is the weighted local feature, and finally the output of the local contextual attention to aspects after extraction by the attention mechanism is shown in Equation (8).

$$O^l = ATT(O_{cdw}) \tag{8}$$

3.3 Sentiment Polarity Classification

3.3.1 Multi-headed Attention Mechanism

The local semantic information and the contextual semantic information are enriched by the multi-headed attention mechanism (MHSA), which can learn different behaviors based on the same attention mechanism, and then combine the different behaviors as knowledge to capture various ranges of dependencies (short-range dependencies and long-range dependencies) within the sequence. The core idea of MHSA is to compute attention values for each word by multiple attention functions, and to perform activation function operations after stitching these multiple attention values together.

$$O^{lg} = O^l \odot O^g \tag{9}$$

$$O^{lg}_{dense} = w^{lg} \cdot O^{lg} + b^{lg} \tag{10}$$

Where O^l and O^g are the local and global features, respectively,$W^{lg} \in R^{d_h \times 2d_h}$, $b^{lg} \in R^{d_h}$ denoting the weights and offsets respectively, and the single-headed attention is calculated as follows.

$$H(O^{lg}_{dense}) = Softmax\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right)V \tag{11}$$

$$Q = O^{lg}_{dense} W^q \tag{12}$$

$$K = O^{lg}_{dense} W^k \tag{13}$$

$$V = O^{lg}_{dense} W^v \tag{14}$$

$$H = H_1 \odot H_2 \odot H_3 \ldots H_h \tag{15}$$

$$Out(O^{lg}_{dense}) = tanh(H \cdot W^{MH}) \tag{16}$$

where $W^q \in R^{d_h \times d_q}$, $W^K \in R^{d_h \times d_k}$, $W^v \in R^{d_h \times d_v}$ are the weight matrices of Q,K,V, respectively; $\odot$denotes the splicing operation of vectors, h is the number of attention heads, and the output of the whole MHSA can be obtained after the calculation of each attention head; is a vector matrix, and Tanh is the activation function. Through the above steps, the multi-head attention encodes the underlying word embedding vector into a semantic vector representation containing rich information.

3.3.2 Output

$$O_{out} = MLP(Out(O_{dense}^{lg})) \tag{17}$$

$$P = softmax(W_p O_{out} + b_p) \tag{18}$$

where MLP denotes multilayer perceptual network, P denotes sentiment polarity classification, W_p and b_p and denotes weights and offsets, respectively.

IV. EXPERIMENTS

4.1. Experimental Datasets And Parameter Configuration

In order to verify the effectiveness of this paper's method for sentiment classification, the datasets for comparison experiments in this paper are two subsets of SEMEVAL2014, Restaurant and Laptop. One is a review set about Restaurant and the other is about Laptop. The datasets contain multiple aspect items and their corresponding sentiment polarity (positive, neutral or negative) with a maximum sentence length of 80. "Fig.2" lists the specific information of the experimental datasets.

TABLE I. EXPERIMENTAL DATASET INFORMATION

Emotion Polarity	Restaurant2014		Laptop2014	
	Training set	*Test set*	*Training set*	*Test set*
positive	994	341	2164	728
neutral	464	169	634	196
negative	870	128	807	196

Fig.2.Experiment datasets

4.1.1 Parameter Configuration

In the experiments of this paper, for the detection of aspect words (the presence or absence of a given aspect in the input sentence, or None if not present), the accuracy Acc and recall F1 values commonly used for classification tasks are used as evaluation metrics in this paper. In this paper, we use a pre-trained BERT for fine-tuning with a feature vector of 12 layers and 768 dimensions. in the model training of this paper, Adam is used as the optimizer with an initial learning rate of 2e- 5, a batch_size of 16, a dropout probability of 0.1, and an epoch of 5.The "Fig.3" lists the value of SRD in each Datasets.

TABLE II. SRD THRESHOLD

Datasets	Restaurant	Laptop
SRD	3	6

Fig.3.Value of SRD

4.2 Experimental results

In this paper, the following classic models are used as baselines to conduct comparative experiments to verify the effectiveness of this method. The evaluation index adopts the accuracy rate and F1 index. The experimental results are shown in "Fig.4".

Compared Methods

(1)ATAE-LSTM[14]:this model is a classical LSTM-based network for the APC task, which applies the attention mechanism to focus on the important words in the context. Besides, ATAE-LSTM appends aspect embedding and the learned features to make full use of the aspect features. The ATAE-LSTM can be adapted to the Chinese review datasets.

(2)IAN[15]: This model proposes a method to interactively learn context and target attention weights.

(3)DCRAN:This model takes the given aspect embedding together with each word embedding as the input of Deep Contextualized Relation-Aware Network, and then embeds the word hidden state into the aspect to generate the attention vector to improve the model's ability to recognize aspect words weights.

(4)BERT-pair[16]: This model proposes a method for auxiliary sentence construction, inputting text and aspect words into BERT in the form of sentence pairs. This method draws on this method.

(5)BERT-LSTM[17]: In this model, all the representation vectors of all intermediate hidden layers in BERT are input into LSTM to increase the depth of text information utilization.

(6)AEN-BERT:this model is an attentional encoder network based on the pretrained BERT model, which aims to solve the aspect polarity classification.

(7)BiLSTM-LCF:Zeng et al. use a semantic-relational distance (SRD) local context-attention design to discard irrelevant sentiment words.

TABLE III. MODEL PERFORMANCE EVALUATION

Model	Restaurant2014		Laptop2014	
	Acc	F1	Acc	F1
ATAE-LSTM	73.04	57.48	68.69	63.17
IAN	78.62	-	72.13	
DCRAN	78.68	64.42	-	-
BERT-pair	81.91	71.90	74.66	68.65
BERT-LSTM	82.22	72.54	75.35	69.34
AEN-BERT	83.12	73.76	79.93	76.31
LCF-BERT	83.84	76.15	79.34	75.61
BiLSTM-LCF	**85.17**	**79.03**	**80.82**	**77.82**

Fig.4.Value of SRD

"Fig.4" shows the comparison results of BiLISTM-LCF with other baselines for Restaurant and Laptop in Semeval2014 Task4 dataset, and it is clear from the experimental results that the aspectual word sentiment classification model designed in this paper can significantly suggest the sentiment classification effect of aspectual words. In this paper, the aspect words and contextual embedding feature vectors are obtained through BERT pre-training, and then the aspect words are modeled separately, and the mask vector designed in this paper is used to weaken the weight of the words that have some influence on the aspect word sentiment polarity but not much influence, so that the perceptual range of the aspect words is strengthened. A global feature extraction module based on BiLSTM is also designed to better capture the global semantic information above to enhance the sentiment expression of aspectual words. Compared with LCF-BERT, the accuracy of this paper's model is improved by 1.33% and 1.48% on Restaurant and Laptop datasets, respectively, and the F1 values are improved by 2.88% and 2.21%, respectively, which proves the effectiveness of this paper's model.

V. CONCLUSION

This paper proposes a fused local and global BiLSTM-LCF model for fine-grained sentiment classification tasks.The aspect-level sentiment classification model in this paper extracts local aspectual features by semantic correlation distance and dynamic masks, and combines the ability of BiLSTM to capture long-distance dependencies to enhance the sentiment representation of local aspectual words, and finally uses a multi-headed attention mechanism to The potential relationship between local aspect words and contextual information is better interacted by using multi-headed attention mechanism. By comparing with other sentiment classification models, the model in this paper improves the accuracy of aspect-level sentiment classification and provides a new idea for the development of ABSA for review texts.

REFERENCES

[1] Pontiki, M.; Galanis, D.; Pavlopoulos, J.; Papageorgiou, H.;Androutsopoulos, I.; and Manandhar, S. 2014. Semeval 2014 task 4: Aspect based sentiment analysis. In SemEval@COLING 2014, 27–35.

[2] Shinhyeok Oh1, Dongyub Lee2*, Taesun Whang et al. Deep Context- and Relation-Aware Learningfor Aspect-based Sentiment Analysis,arXiv:2016.038.6v1.2021.

[3] Zeng B, Yang H, Xu R, et al. Lcf: A local context focus mechanism for aspect-based sentiment classification[J]. Applied Sciences, 2019, 9(16): 3389.

[4] Wu Z, Ong D C. Context-guided bert for targeted aspect-based sentiment analysis[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Online, Feb 2-9, 2021, 35(16): 14094- 14102. Hochreiter S，Schmidhuber J.Long short-term memory［J］.Neural Computation，1997，9（8）：1735-1780.

[5] WU B R, QIAO H, JIA Y F, et al. Sentiment Analysis of Mid length Microblog Based on Capsule Network[J]. Journal of Signal Processing,2022,38(8):1632-1641. DOI: 10.16798/j. issn.1003-0530.2022.08.008.

[6] TANG D Y, QIN B, FENG X C, et al. Effective LSTMs for target-dependent sentiment classification[C]//Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers, 2016: 3298-3307

[7] 陈虹, 杨燕, 杜圣东. 用户评论方面级情感分析研究[J]. 计算机科学与探索, 2021, 15(3): 478-485.

[8] 徐志栋, 陈炳阳, 王晓, 等. 基于胶囊网络的方面级情感分类研究[J]. 智能科学与技术学报, 2020, 2(3): 284-292

[9] 宋婷, 陈战伟, 杨海峰. 基于分层注意力网络的方面情感分析[J]. 大数据, 2020, 6(5): 82-91.

[10] Peng, H., Ma, Y., Li, Y., Cambria, E., 2018. Learning multi-grained aspect target sequence for chinese sentiment analysis. Knowledge-Based Systems 148, 167–176. doi:10.1016/j.knosys.2018.02.034.

[11] Chen, F., Huang, Y., 2019. Knowledge-enhanced neural networks for sentiment analysis of chinese reviews. Neurocomputing 368, 51 – 58.

[12] Liu, N., Shen, B., 2019. Aspect-based sentiment analysis with gated alternate neural network. Knowledge-Based Systems , 105010.

[13] Zeng, Z., Ma, J., Chen, M., Li, X., 2019b. Joint learning for aspect category detection and sentiment analysis in chinese reviews, in: China Conference on Information Retrieval, Springer. pp. 108–120.

[14] WANG Y Q, HUANG M L, ZHU X Y, et al. Attention-based LSTM for aspect-level sentiment classification[C]//Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2016: 606-615.

[15] Ma, D., Li, S., Zhang, X. & Wang, H. Interactive attention networks for aspect-level sentiment classification[C]. Proceedings of the twenty-sixth international joint conference on artificial intelligence.2017:4069–4074.

[16] Chi Sun, Luyao Huang, Xipeng Qiu*. Utilizing BERT for Aspect-Based Sentiment Analysis via Constructing Auxiliary Sentence.[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics :Human Language Technologies,2019:380-385

[17] LI Pan,WU Yadong,CHU QikaiFU,Chaoshuai,ZHANG Guiyu,et al.Aspect-based sentiment analysis of long text based on BERT and memory network [J]. Transducer and Microsystem Technologies,2022(2):118-122.