Learnable Conjunction Enhanced Model for Chinese Sentiment Analysis

Bingfei Zhao, Hongying Zan *, Jiajia Wang, Yingjie Han
Zhengzhou University
zbf472670570@163.com, iehyzan@zzu.edu.cn, wjj4work@163.com
ieyjhan@zzu.edu.cn

Abstract

Sentiment analysis is a crucial text classification task that aims to extract, process, and analyze opinions, sentiments, and subjectivity within texts. In current research on Chinese text, sentence and aspect-based sentiment analysis is mainly tackled through well-designed models. However, despite the importance of word order and function words as essential means of semantic expression in Chinese, they are often underutilized. This paper presents a new Chinese sentiment analysis method that utilizes a Learnable Conjunctions Enhanced Model (LCEM). The LCEM adjusts the general structure of the pre-trained language model and incorporates conjunctions location information into the model's fine-tuning process. Additionally, we discuss a variant structure of residual connections to construct a residual structure that can learn critical information in the text and optimize it during training. We perform experiments on the public datasets and demonstrate that our approach enhances performance on both sentence and aspect-based sentiment analysis datasets compared to the baseline pre-trained language models. These results confirm the effectiveness of our proposed method.

1 Introduction

Sentiment analysis is a crucial area of research within the field of natural language processing. Before the advent of Transformer (Vaswani et al., 2017), Recurrent Neural Networks (RNNs) were the primary method used to model sequences in language modeling tasks (Tang et al., 2016a; Li et al., 2018; Li et al., 2019; Majumder et al., 2022). RNN, along with its variants LSTM (Long-Short Term Memory) and GRU (Gated Recurrent Unit), are powerful models for processing sequences of varying lengths and addressing long-term dependencies. However, the sequential nature of RNNs makes parallelization difficult. Transformer introduces the attention mechanism to encode the context information, which can well capture the internal correlation and ease the problem of long-term dependencies. This allows for greater parallelization and improved performance on certain tasks.

Nevertheless, since self-attention discards sequential operations when processing sequences, the position information in the sequence cannot be fully utilized. In languages such as Chinese, word order plays a crucial role in conveying grammatical meaning [0], making it important to consider the sequential nature of the language when developing natural language processing models. Word order refers to the sequence of words in a phrase or sentence, while Chinese word order is relatively fixed, and the change of word order can make the phrase or sentence express different meanings. "Speak well/说好话", "easy to speak with/好说话", and "easier said/话好说" are three Chinese phrases that demonstrate the importance of word order in conveying meaning. Although these phrases share similar characters, their meanings differ greatly depending on how those characters are arranged. "Speak well/说好话" means to speak positively or say good things about someone or something, while "easy to speak with/好说话" describes someone who is easy to communicate with. Lastly, "easier said/话好说" implies that something may sound simple or easy to do but can be more difficult in practice. It's essential to consider

*Corresponding author

[0]Higher Education Press.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

both the context and word order when interpreting or translating Chinese phrases. In addition, function words in Chinese play an important role in constructing the grammatical structure of a sentence and reflecting specific grammatical relationships. They are a crucial grammatical tool necessary for expressing meaning [1]. Among them, conjunctions connect grammatical units at different levels, and their positions in sentences are significantly different (Liu, 2016), which can be used as an essential aspect of studying syntactic distribution.

Therefore, in this paper, we propose LCEM, a learnable conjunctions augmentation model for Chinese sentiment analysis. By adjusting the structure of the pre-trained language model, LCEM introduces the conjunction position information into the fine-tuning process. The paper also explores variants of residual structure and constructs an enhanced model capable of learning critical information during training and optimization of the residual structure.

The main contributions of this paper can be summarized as follows:

- LCEM is a generic structure that can be easily integrated into a pre-trained language model based on Transformer using an adaptive update optimized network of learnable parameter factors.
- By incorporating the relative position of conjunctions in each layer of the pre-trained language model, LCEM enhances multi-head self-attention and effectively considers the sentiment range of sentences connected by conjunctions.
- Additionally, LCEM combines a learnable residual structure to better balance the network and optimize semantic representation more efficiently.
- LCEM is evaluated on benchmark datasets for sentence and aspect-based sentiment analysis. Experiments show that LCEM consistently achieves state-of-the-art performance across all test datasets.

2 Related Work

2.1 Chinese sentiment Analysis

Early Chinese sentiment analysis methods (Zhu et al., 2006; SHI Wei, 2021; Liu et al., 2015) primarily relied on sentiment lexicons, such as HowNet sentiment word dictionary and National Taiwan University Sentiment Dictionary (NTUSD), and classified sentiment polarity based on dictionaries and rules. However, these methods are limited by the quality and coverage of lexicons. The sentiment analysis in a specific field needs to construct a specific dictionary, which is time-consuming and laborious. When traditional machine learning algorithms are used in sentiment classification, different features enable different classifiers to obtain higher accuracy than dictionary methods (Xu et al., 2007; Yang and Lin, 2011; He et al., 2018). However, traditional machine learning methods rely on the quality of the annotated corpus and cannot fully use contextual semantic information.

With the rapid development of deep learning, neural network and attention mechanism have been widely concerned and applied in Chinese sentiment analysis (Cheng et al., 2019; Peng et al., 2018). Transformer with self-attention mechanism, which employs an encoder-decoder framework to better address long-term dependencies and allows for more robust scalability of parallel computations, is widely used in natural language processing. Based on the Transformer architecture, a series of landmark pre-trained language models have emerged, showing a strong ability to learn generic Chinese representations. Li (2021) fully extracted context information using improved attention to encode relative position between words based on ELMo (Peters et al., 2018). Xie(2020) used BERT to encode the set of sentiment words extracted from texts and used attention to obtain sentiment information. However, in the above studies, although the pre-trained language model has powerful modeling ability, it neglects the application of syntactic structure or semantic information in sentiment analysis and fails to use sentiment features effectively.

[1]The Commercial Press.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

2.2 Relative Position Feature

In order to leverage the sequential information contained within input text, Transformers incorporate position embeddings into the original input embedding. This process is calculated as follows:

$$
\begin{aligned}
PE_{(pos,2i)} &= sin(pos/10000^{(2i/d_{model})}) \\
PE_{(pos,2i+1)} &= cos(pos/10000^{(2i/d_{model})})
\end{aligned}
\tag{1}
$$

where pos represents position, i represents the number of dimensions, d_{model} is the input and output vector dimensions. The sines and cosines enable the model to learn the relative position and easily extend to longer sequences.

The BERT-based pre-trained language model adopts the encoder structure in Transformer and selects absolute position embedding to better adapt to downstream tasks. In the input layer, word embedding is combined with position embedding to ensure that identical words at different positions can learn representations that are appropriate for their respective contexts. Li (2021) improved attention by encoding relative positions between words. Shaw (2018) used relative encoding as an additional value in the self-attention to capture information about the relative position differences between input elements. According to different task characteristics, different position embeddings contain different meanings. For instance, in the named entity recognition task, entity term is often introduced by designing different position features (Li et al., 2020; Yan et al., 2019; Mengge et al., 2020). In the causality extraction task, position features can reflect the position of connectives and the distance between causal events and connectives (Zhao et al., 2016).

2.3 Residual Structure

Neural networks have a strong representation ability and can optimize and update the network structure through the back propagation algorithm. However, during backpropagation, gradients may either vanish or increase exponentially, resulting in ineffective updates to the underlying parameters, or gradient explosion. Furthermore, deeper networks are susceptible to degradation problems. He (2016) verified that adding more layers to a network model with a certain depth will lead to higher training errors.

Recently, residual learning has been widely used in natural language processing and computer vision as a technique for optimization of deep neural network to alleviate gradient vanishing or explosion problems (He et al., 2016; Srivastava et al., 2015; Liu et al., 2019a; Liu et al., 2021). Since each submodule of the Transformer encoder contains residual structures with layer normalization, BERT-based pre-trained variants can also make full use of residual connections to optimize the network.

This paper introduces the learnable residual structure based on enhanced self-attention by the position features of conjunctions. By assigning learnable parameters to each branch, the residual structure can be adjusted adaptively, and performance can be improved through simple model adjustment.

3 Methodology

3.1 Overview

LCEM is based on the basic architecture of the pre-trained language model. The overall structure of LCEM is described in Figure 1. LCEM uses the conjunction relative position enhanced multi-head attention to replace the multi-head attention module in each layer of the pre-trained language model. By combining the relative position feature with the attention mechanism, the model can learn global semantic information while still paying close attention to important local ranges. In addition, the residual structure of the pre-trained language model is improved to a more flexible structure to optimize the network and enable better internal information sharing. The learnable factors can adaptively control the residual structure, better integrating the semantic information learned by the relative position feature and further optimizing by assigning different importance to each residual branch.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

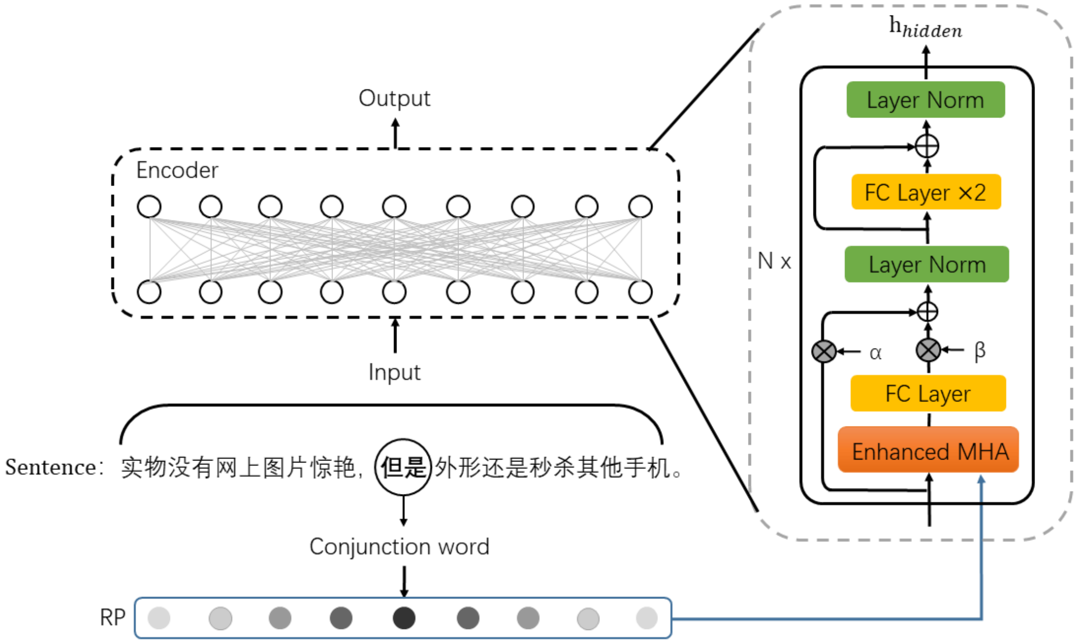

Figure 1: Overview of LCEM

3.2 Conjunction Relative Position Enhanced Multi-Head Attention

LCEM uses the relative position feature to enhance attention to learn the interaction between input text and the conjunctions representation. Conjunctions of transition, progression, selection, and coordinate are selected in the Chinese Function Word Usage Knowledge Base(CFKB) (Zan et al., 2011; Kunli et al., ; Zhang et al., 2015), and the distance $d(d \geq 0)$ between each character in a sentence and the first character of the conjunction is calculated. We map the relative position of conjunctions into the interval of $(0, 1)$ to obtain the relative position feature RP, and the calculation is as follows:

$$RP = 1 - Sigmoid(d) = 1 - \frac{1}{1 + e^{-d}} \tag{2}$$

If there is no conjunctions in the sentences, the d in the formula is the distance between each word in the sentences and the beginning of the sentences.

Then, as shown in Figure 2, RP increases the attention to the context near conjunctions. At the same time, the learnable parameter ω is introduced to reduce the noise caused by introducing the relative position feature to the original input representation H. The attention after adding the relative position feature is as follows:

$$\begin{aligned} Attetnion(Q, K, V) &= Softmax(\frac{QK^T}{\sqrt{d_K}} + \omega RP)V \\ where \quad Q &= HW^Q, K = HW^K, V = HW^V \end{aligned} \tag{3}$$

3.3 Learnable Residual Structure

Some studies (Liu et al., 2019a; Liu et al., 2021) divided the problems existing in residual connection into two types: the balance problem of each residual branch and the optimization problem. Liu (2019a) analyzed existing works and summarized the general residual structure as follows:

$$\mathcal{Y} = \alpha x + \beta \mathcal{F} + \gamma LN(x + \mathcal{F}) \tag{4}$$

Where x is the input branch, i.e., the skip connection, $\mathcal{F}$ is the residual branch, LN is layer normalization, $\mathcal{Y}$ is the output of the residual block, and α, β, γ are the weight factors. The residual block can be

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

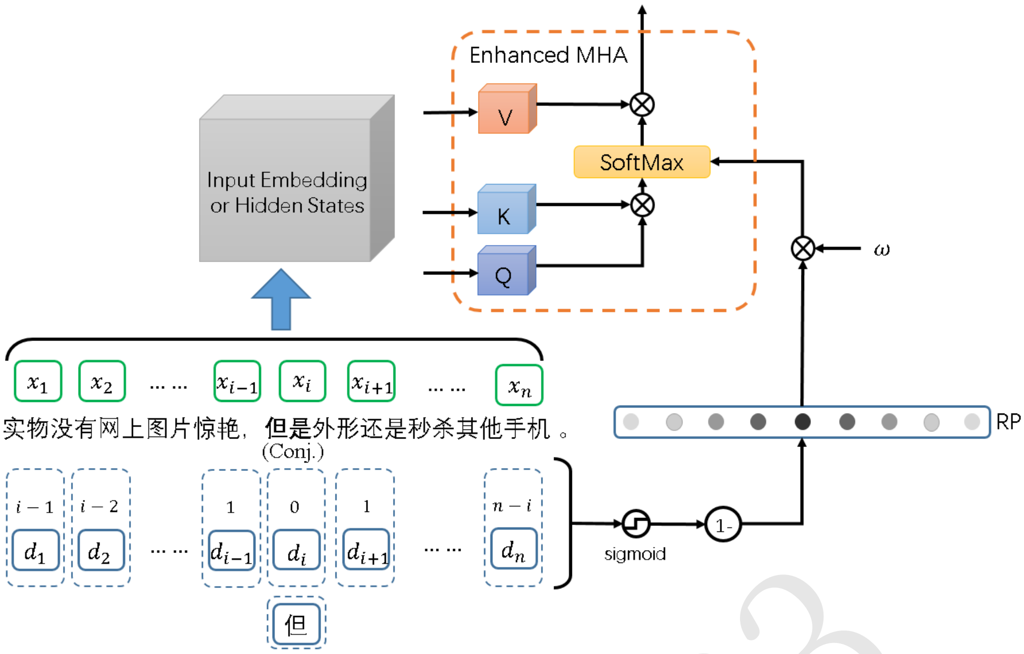

Figure 2: Details of Conjunction Relative Position Enhanced Multi-Head Attention

(a) Transformer (b) Learnable Residual Structure in LCEM

Figure 3: Residual Structure in Transformer and LCEM

adjusted and optimized adaptively by adjusting values for α, β, and γ. Liu (2021) proposed formula 5 to summarize the residual connection with normalization. Normalization $\mathcal{G}$ was placed outside the sum of input x and nonlinear transformation $\mathcal{F}(x, W)$, and λ was used to enhance the input branch.

$$\mathcal{Y} = \mathcal{G}(\lambda x + \mathcal{F}(x, W)) \tag{5}$$

Drawing inspiration from the residual structure present in every layer of the Transformer (Figure 3 (a)), layer normalization plays a crucial role in the model's overall performance. It can help the optimization of nonlinear transformation to a certain extent. And, in combination with the idea of adjusting each branch of residual in the neural network by the weight factor mentioned above, the residual structure is summarized as follows:

$$\mathcal{Y} = LN(\alpha x + \beta \mathcal{F}) \tag{6}$$

As shown in Figure 3 (b), the residual structure in Transformer can be regarded as a particular case $\mathcal{Y} = LN(x + \mathcal{F})$ when $\alpha = \beta = 1$. In Transformer, the residual branch $\mathcal{F}$ can be either multi-headed attention or feedforward networks. In this paper, we focus on the residual structure of multi-head attention. We propose to replace the residual branch with conjunctions relative position enhanced

Table 1: Statistical data of each category in the datasets.

Datasets	COAE2013		NLPCC2014		SemEval16_CAM		SemEval16_PHO	
	Train	Test	Train	Test	Train	Test	Train	Test
Positive	753	305	5000	1250	809	344	758	310
Negative	876	239	5000	1250	450	137	575	219

attention. Meanwhile, α and β are set as learnable parameters so that the model can self-learn appropriate scaling factors. The proportion of input branch x and residual branch $\mathcal{F}$ in the network is constantly modified to achieve optimization.

The semantic representation obtained by the enhanced attention will further learn the appropriate proportion in the propagation under the adjustment of scaling factor β, reducing the noise caused by the introduction of the relative position feature. Scaling factors α and β jointly determine the different distribution of x and $\mathcal{F}$. The layer normalization is used to make the distribution of each layer in the network relatively consistent to avoid gradient vanishing or explosion caused by the change of learnable parameters. Through multi-layer structure with learnable conjunctions enhanced attention, the final output is obtained by a linear classifier.

4 Experimental Settings

4.1 Datasets

In this paper, we study two granular subtasks in Chinese sentiment analysis. Statistical data of the above datasets are shown in Table 1.

For Chinese sentence-level sentiment analysis, COAE2013 and NLPCC2014 are selected. COAE2013 is a dataset of annotated data from The Fifth Chinese Opinion Analysis Evaluation, consisting of 1004 positive reviews and 834 negative reviews. The dataset was divided into train set and test set according to the ratio of 9:1. NLPCC2014 is from the 3rd CCF Conference on Natural Language Processing & Chinese Computing, including reviews of books, DVDs, electronic products, and other domains. The train set consisted of 5,000 positive and 5,000 negative texts, and the test set consisted of 2,500 texts.

For the Chinese aspect-based sentiment analysis task, this paper selects SemEval2016 (Pontiki et al., 2016). Task 5 of SemEval2016 provides a Chinese dataset of electronic product aspect-based reviews in two specific domains, including phone and camera, including 400 samples, a total of about 4100 sentences.

4.2 Baselines

We evaluate LCEM with typical sentiment analysis and text classification models as baselines for sentence-level sentiment analysis, including BiLSTM (Zhang et al., 2015), BiLSTM+Att (Zhang and Wang, 2015), TextCNN (Kim, 2014), DPCNN (Johnson and Zhang, 2017), and pre-trained language models like EBi-SAN (2021), BERT, BERT_wwm (Cui et al., 2021), RoBERTa (Liu et al., 2019b), ERNIE (Sun et al., 2019b). For aspect-based sentiment analysis, we compare our solution to several models that can be applied to Chinese text, including MemNet (Tang et al., 2016b), ATAE-LSTM (Wang et al., 2016), IAN (Ma et al., 2017), Ram (Chen et al., 2017), AOA (Huang et al., 2018), MGAN (Fan et al., 2018), Tnet (Li et al., 2018), and QA-B (Sun et al., 2019a) and NLI-B (Sun et al., 2019a), and also BERT and ERNIE.

The word vector pre-trained by the Sogou News corpus is selected as the initial embedding in the general baselines. The batch size is 128, the learning rate is 1E-5, and 30 epochs are trained by Adam optimization. Based on the pre-trained model, the baselines all follow the default 12 hidden layers with a size of 768, the batch size is 20, and the learning rate is 5E-5. Adam is used to optimize the cross-entropy loss function and fine-tunes the parameters.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.
(c) Technical Committee on Computational Linguistics, Chinese Information Processing Society of China

5 Experimental Results

5.1 Results on Sentence-level Sentiment Analysis

Table 2 shows the results of comparative experiments on the sentence-level datasets.Compared with the pre-trained model ERNIE and neural network models based on RNN and CNN, such as TextCNN and DPCNN, the results indicate that the fine-tuned pre-trained language model performs better on the datasets than the neural network models based on RNN and CNN, highlighting the huge advantage of pre-trained language models in sentiment analysis tasks. Additionally, compared to other pre-trained models, ERNIE performs better on two sentiment analysis datasets. By using relative positional encoding of conjunctions and learnable residual structures based on ERNIE, LCEM further optimized the model and improved its performance, demonstrating the effectiveness of the proposed method in this paper.

Table 2: Results on sentence-level sentiment analysis datasets.

Datasets	COAE2013		NLPCC2014	
	Acc(%)	F1(%)	Acc(%)	F1(%)
BiLSTM	85.74	85.39	60.48	60.48
BiLSTM+Att	86.91	86.76	69.60	69.56
TextCNN	89.65	89.46	69.04	68.85
DPCNN	87.30	87.07	62.48	58.88
EBi-SAN	-	-	79.08	78.48
BERT	93.57	93.53	79.61	79.61
BERT_wwm	94.88	94.83	80.21	80.20
RoBERTa	94.99	95.01	79.57	79.56
ERNIE	95.77	95.74	80.89	80.88
LCEM	**96.69**	**96.68**	**81.08**	**81.08**

5.2 Results on Aspect-based Sentiment Analysis

Experimental results are shown in Table 3 compared with aspect-based sentiment analysis baselines. Under the accuracy and F1, LCEM outperforms all baselines in SemEval16_CAM and SemEval16_PHO. The accuracy of LCEM on the SemEval16_CAM is 1.25% higher than that of ERNIE, and the F1 value is 0.72% higher than that of QA-B. Compared with IAN, MGCN, and other non-pre-trained language models, the fine-tuned results of the pre-trained model have great advantages. On the one hand, the pre-trained model has been trained on large text corpus and has learned rich language representation capabilities, which enables the pre-trained model to better understand the semantics and context of the text, which is very helpful for sentiment analysis tasks. On the other hand, pre-trained models can achieve better results on small datasets, while recurrent neural networks require large amounts of manually annotated training data, and the size of the training data will limit the performance of the model.

5.3 Ablation Study

Table 4 shows the results of LCEM ablation experiments on four datasets.

In which, $+RP$ and $+\omega RP$ respectively represent adding relative position encoding (RP) and weighted relative position encoding (Weighted RP) only in the self-attention module on top of the baseline model. Comparing $+RP$ and $+\omega RP$ with baseline ERNIE, we can see that $+\omega RP$ is better than $+RP$, improves performance on both sentence-level datasets and SemEval16_PHO. But on SemEval16_CAM, neither $+RP$ nor $+\omega RP$ can achieve effective performance enhancement, which may be because the relative position feature is added to each layer of the pre-trained language model. The output of each layer will serve as input to the next layer and participate in the residual structure. As the network depth increases, each addition of the relative position feature will introduce some noise into the original representation. Although the weighted relative position feature($+\omega RP$) introduces parameters that can learn relative positional shifts with the network structure, its effect varies on different datasets.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

Table 3: Results on aspect-based sentiment analysis datasets.

Datasets		SemEval16_CAM		SemEval16_PHO	
		Acc(%)	F1(%)	Acc(%)	F1(%)
ATAE-LSTM		87.11	82.79	79.02	78.78
MemNet		88.57	85.33	77.88	76.77
IAN		88.77	85.97	79.40	78.91
Ram		85.65	82.66	77.69	76.81
Tnet		87.32	83.47	79.77	79.14
AOA		88.36	85.52	79.58	79.21
MGAN		85.45	82.65	79.96	79.38
BERT		87.94	85.57	83.74	83.22
ERNIE		93.14	91.45	90.17	89.84
ERNIE-SPC		92.52	90.65	90.36	90.07
ERNIE-based	QA-B	92.41	92.41	89.23	89.22
	NLI-B	91.48	91.48	88.94	88.94
LCEM		**94.39**	**93.13**	**91.12**	**90.79**

Table 4: Results of ablation experiment

Datasets	COAE2013		NLPCC2014		SemEval16_CAM		SemEval16_PHO	
	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)	Acc(%)	F1(%)
Baseline(ERNIE)	95.77	95.74	80.89	80.88	93.14	91.45	90.17	89.84
$+RP$	95.96	95.94	80.76	80.75	92.93	91.77	89.60	89.17
$+\omega RP$	95.96	95.93	80.92	80.91	92.93	91.32	90.74	90.42
$+LRS$	96.51	96.49	80.40	80.40	92.31	90.74	90.17	89.83
$+RP\&LRS$	95.96	95.93	80.96	80.95	93.35	**93.35**	90.55	90.27
$+\omega RP\&LRS$(LCEM)	**96.69**	**96.68**	**81.08**	**81.08**	**94.39**	93.13	**91.12**	**90.79**

$+LRS$ represents only the learnable residual structure added to ERNIE. The comparison results also show that $+LRS$ has a slight improvement, indicating that the structure of the pre-trained language model, especially the residual structure, has the advantages of efficiency, stability, and universality.

Accuracy and macro-F1 of $+RP\&LRS$ are better than $+RP$, $+\omega RP$, and $+LRS$ in both datasets. This suggests that scaling within the residual structure can effectively adjust the enhanced multi-head attention as a branch of residual connection. In addition, the output of the previous layer serves as the skip connection branch of the residual structure of the next layer, and residual scaling can adjust the input branch and the residual branch adaptively. At the same time, it shows that enhanced attention by relative location features can capture both content and distance information, and learn richer context representation under the role of location information.

The proposed model LCEM($+\omega RP\&LRS$) achieves the highest accuracy and F1 in both sentence-level datasets. In the two datasets of SemEval16, the F1 improved by 1.68% and 0.95%, respectively, compared with baseline model ERNIE, and achieved the highest accuracy in both datasets. Compared with $+RP\&LRS$, the accuracy is significantly improved, indicating that weighted relative position encoding can achieve more effective optimization. The learnable weights during network training also reduce the noise effects introduced by relative position encoding, better capture the balance within the network and maximizing the gain of residual scaling.

5.4 Case Study

For further analysis of the model, the LCEM and ERNIE models are analyzed in this paper, as shown in Table 5.

For the adversative conjunction ”但是”, it serves as a transitional element between two sentences or clauses. It indicates a contrast or contradiction between the information presented before and after it. In the given context, the emotional tone of the sentence preceding the transition is predominantly negative. However, the emotional tone of the sentence following the use of ”但是” changes from negative to

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

Table 5: Case studies of LCEM and ERNIE models

Type of conjunction	Conjunction	Example	Model	Label
转折	但是	拿到的时候还觉得像盗版，但确实是正版的，很完整，非常不错	ERNIE	0
			LCEM	1
递进	而且	是真正的职场小说，感觉更像《圈子圈套》，而且厚厚的一大本，很值。	ERNIE	0
			LCEM	1

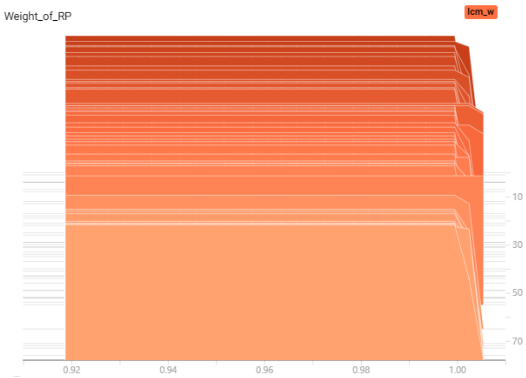

Figure 4: The parameter ω of RP over time.

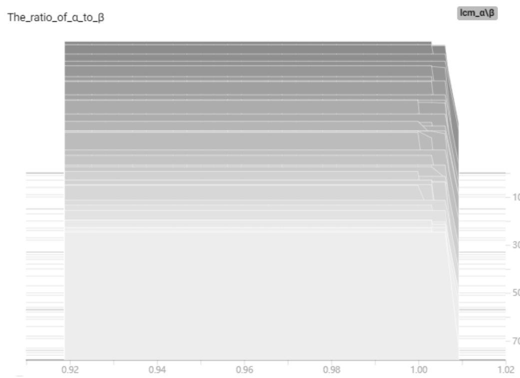

Figure 5: The ratio of α to β over time.

positive. Therefore, the emotional label of the first sentence in Table 5 is 1, signifying a shift from negative to positive emotion.

On the other hand, the coordinating conjunction "而且" is used to connect two sentences or clauses to express a progression or addition of information. While the emotional information in the sentence before the conjunction may not be overtly expressed, it is more fully conveyed in the sentence that follows the use of "而且." Consequently, the emotional label of the second sentence in Table 5 is 1, indicating the enhanced expression of emotional content instead of label 0. When compared to ERNIE, LCEM, which incorporates conjunctive information, provides more accurate predictions of emotional labels.

6 Learnable Parameters Analysis

Figure 4 and Figure 5 show the changes of relative position parameter ω and α to β ratio over time. The X-axis represents the range of parameter values, while the Y-axis on the right represents the number of training steps. Each slice in the figure is a single histogram, representing the distribution of parameters in a training step. The number of training steps is gradually increased from back to front.

According to Figure 4, the learnable parameter ω of the relative position feature RP is more evenly distributed in $[0.919, 0.999]$, indicating that the relative position feature occupies a vital proportion of attention. Moreover, combined with the ablation experiment results in Section 5.4, relative location feature enhanced attention can capture both content and distance information and learn a richer context representation under the effect of location information.

Figure 5 shows that the ratio of α to β is evenly distributed in $[0.919, 1.01]$. In most cases, the proportion of input branches is smaller than that of residual branches. In each Transformer encoder, the proportion of representations from the previous layer is smaller than that of expressions enhanced by the relative position of the conjunctions. It demonstrates the significance of the semantic representation obtained through enhanced attention in the network. Moreover, input branches also play an important role in network. Through layer-by-layer propagation, the semantic representation acquired by each layer can be preserved in the lower layers and will participate in the attention mechanism to further extract abstract semantics. The learnable parameters greatly help the information transfer and optimization of network structure.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

7 Conclusion

In this paper, we introduce LCEM, a model that incorporates semantic information using relative position features of conjunctions, and guides the Chinese sentiment analysis task through adaptive residual structure.Specifically, weighted relative position features reduce the introduced noise and improve the learning ability of location-related syntactic features, which can better guide the self-attention mechanism and help the model focus on the critical sentences for semantic representation. At the same time, we propose a novel learnable residual structure based on pre-trained language models that can effectively handle the interaction between residual and input branches in an adaptive manner. Experimental results show that our method is effective in Chinese sentiment analysis, where relative position and adaptive residual structure complement each other. The relative position information helps the model to focus on crucial information for sentiment analysis, while the residual structure in each layer balances the learned knowledge within the network structure.

Acknowledgements

This research is supported by the key research and development and promotion project of Henan Provincial Department of Science and Technology in 2023: Research on Automatic Question Answering System in Science and Technology Management Based on Knowledge Graph (232102211041) and the National Social Science Foundation project: Research on Knowledge Base and Application of Modern Chinese Function Word Usage for Natural Language Processing (14BYY096). The author would like to thank the anonymous reviewers for their valuable comments and suggestions on the improvement of this paper.

References

Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang. 2017. Recurrent attention network on memory for aspect sentiment analysis. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 452–461.

Yan Cheng, Z Ye, M Wang, Q Zhang, and G Zhang. 2019. Chinese text sentiment orientation analysis based on convolution neural network and hierarchical attention network. *Journal of Chinese Information Processing*, 33(01):133–142.

Yiming Cui, Wanxiang Che, Ting Liu, Bing Qin, and Ziqing Yang. 2021. Pre-training with whole word masking for chinese bert. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3504–3514.

Feifan Fan, Yansong Feng, and Dongyan Zhao. 2018. Multi-grained attention network for aspect-level sentiment classification. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 3433–3442.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Y He, S Zhao, and L He. 2018. Micro-text emotional tendentious classification based on combination of emotion knowledge and machine-learning algorithm. *J. Intell*, 37(5):189–194.

Binxuan Huang, Yanglan Ou, and Kathleen M. Carley. 2018. Aspect level sentiment classification with attention-over-attention neural networks. In Robert Thomson, Christopher Dancy, Ayaz Hyder, and Halil Bisgin, editors, *Social, Cultural, and Behavioral Modeling*, pages 197–206, Cham. Springer International Publishing.

Rie Johnson and Tong Zhang. 2017. Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 562–570.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar, October. Association for Computational Linguistics.

Zhang Kunli, Zan Hongying, Chai Yumei, Han Yingjie, and Zhao Dan. Construction and application of the chinese function word usage knowledge base.

Xin Li, Lidong Bing, Wai Lam, and Bei Shi. 2018. Transformation networks for target-oriented sentiment classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 946–956, Melbourne, Australia, July. Association for Computational Linguistics.

Xin Li, Lidong Bing, Piji Li, and Wai Lam. 2019. A unified model for opinion target extraction and target sentiment prediction. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6714–6721.

Xiaonan Li, Hang Yan, Xipeng Qiu, and Xuanjing Huang. 2020. FLAT: Chinese NER using flat-lattice transformer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6836–6842, Online, July. Association for Computational Linguistics.

Z Li, L Chen, and S Zhang. 2021. Chinese text sentiment analysis based on elmo and bi-san. *Application Research of Computers*, 38(8):2301–2307.

YJ Liu, SG Ju, SM Wu, and C Su. 2015. Classification of chinese texts sentiment based on semantic and conjunction. *Journal of Sichuan University Natural Science Edition*, 52:57–62.

Fenglin Liu, Meng Gao, Yuanxin Liu, and Kai Lei. 2019a. Self-adaptive scaling approach for learnable residual structure. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 862–870.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Fenglin Liu, Xuancheng Ren, Zhiyuan Zhang, Xu Sun, and Yuexian Zou. 2021. Rethinking skip connection with layer normalization in transformers and resnets. *arXiv preprint arXiv:2105.07205*.

Qun Liu. 2016. Study of conjunctive scope and its "special category" in modern chinese. *Journal of Jiangsu Normal University: Philosophy Social Science Edtion*, 42(85-90).

Dehong Ma, Sujian Li, Xiaodong Zhang, and Houfeng Wang. 2017. Interactive attention networks for aspect-level sentiment classification. AAAI Press.

Navonil Majumder, Rishabh Bhardwaj, Soujanya Poria, Alexander Gelbukh, and Amir Hussain. 2022. Improving aspect-level sentiment analysis with aspectnbsp;extraction. *Neural Comput. Appl.*, 34(11):8333–8343, jun.

Xue Mengge, Bowen Yu, Tingwen Liu, Yue Zhang, Erli Meng, and Bin Wang. 2020. Porous lattice transformer encoder for Chinese NER. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3831–3841, Barcelona, Spain (Online), December. International Committee on Computational Linguistics.

Haiyun Peng, Yukun Ma, Yang Li, and Erik Cambria. 2018. Learning multi-grained aspect target sequence for chinese sentiment analysis. *Knowledge-Based Systems*, 148:167–176.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June. Association for Computational Linguistics.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Véronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nuria Bel, Salud María Jiménez-Zafra, and Gülşen Eryiğit. 2016. SemEval-2016 task 5: Aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*, pages 19–30, San Diego, California, June. Association for Computational Linguistics.

Peter Shaw, Jakob Uszkoreit, and Ashish Vaswani. 2018. Self-attention with relative position representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 464–468, New Orleans, Louisiana, June. Association for Computational Linguistics.

FU Yue SHI Wei. 2021. Microblog short text mining considering context:a method of sentiment analysis. *Computer Science*, 48(6A):158.

Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway networks. *arXiv preprint arXiv:1505.00387*.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.

Chi Sun, Luyao Huang, and Xipeng Qiu. 2019a. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 380–385, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019b. Ernie: Enhanced representation through knowledge integration. *arXiv preprint arXiv:1904.09223*.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016a. Effective LSTMs for target-dependent sentiment classification. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 3298–3307, Osaka, Japan, December. The COLING 2016 Organizing Committee.

Duyu Tang, Bing Qin, and Ting Liu. 2016b. Aspect level sentiment classification with deep memory network. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 214–224, Austin, Texas, November. Association for Computational Linguistics.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yequan Wang, Minlie Huang, Xiaoyan Zhu, and Li Zhao. 2016. Attention-based lstm for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pages 606–615.

R Xie and Y Li. 2020. Text sentiment classification model based on bert and dual channel attention. *Journal of Data Acquisition and Processing*, 35(4):642–652.

Jun Xu, Yu-Xin Ding, and Xiao-Long Wang. 2007. Sentiment classification for chinese news using machine learning methods. *Journal of Chinese Information Processing*, 21(6):95–100.

Hang Yan, Bocao Deng, Xiaonan Li, and Xipeng Qiu. 2019. Tener: adapting transformer encoder for named entity recognition. *arXiv preprint arXiv:1911.04474*.

Jing Yang and Shiping Lin. 2011. Emotion analysis on text words and sentences based on svm. *Jisuanji Yingyong yu Ruanjian*, 28(9):225–228.

Hongying Zan, Kunli Zhang, Xuefeng Zhu, and Shiwen Yu. 2011. Research on the chinese function word usage knowledge base. *Int. J. Asian Lang. Process.*, 21(4):185–198.

Dongxu Zhang and Dong Wang. 2015. Relation classification via recurrent neural network. *arXiv preprint arXiv:1508.01006*.

KL Zhang, HY Zan, YM Chai, Yingjie Han, and Dan Zhao. 2015. Survey of the chinese function word usage knowledge base. *Journal of Chinese Information Processing*, 29(3):1–8.

Sendong Zhao, Ting Liu, Sicheng Zhao, Yiheng Chen, and Jian-Yun Nie. 2016. Event causality extraction based on connectives analysis. *Neurocomputing*, 173:1943–1950.

Yan-Lan Zhu, Jin Min, Ya-qian Zhou, Xuan-jing Huang, and Li-De Wu. 2006. Semantic orientation computing based on hownet. *Journal of Chinese information processing*, 20(1):14–20.

Proceedings of the 22nd China National Conference on Computational Linguistics, pages 761-772, Harbin, China, August 3 - 5, 2023.