

Retraction

Retracted: Chinese Short Text Classification by ERNIE Based on LTC_Block

Wireless Communications and Mobile Computing

Received 27 June 2023; Accepted 27 June 2023; Published 28 June 2023

Copyright © 2023 Wireless Communications and Mobile Computing. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] M. Zhang and X. Shang, "Chinese Short Text Classification by ERNIE Based on LTC_Block," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1411744, 8 pages, 2022.

Research Article

Chinese Short Text Classification by ERNIE Based on LTC_Block

Meng Zhang  and Xiang Shang

Department of Statistics, Tianjin University of Finance and Economics, Tianjin 300222, China

Correspondence should be addressed to Meng Zhang; zhangmeng@tjufe.edu.cn

Received 29 December 2021; Accepted 22 February 2022; Published 9 April 2022

Academic Editor: Mohammad Farukh Hashmi

Copyright © 2022 Meng Zhang and Xiang Shang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Short text classification, an important direction of the basic research of natural language processing, has extensive applications. Its effect depends on feature extraction methods and feature representation methods. This paper proposed an LTC_Block-based short text classification model named ERNIE to classify Chinese short texts and extract semantics in the corpus to address the polysemy problem in the text. In this model, LTC_Block, a double-channel structural unit composed of BiLSTM and TextCNN, was used to extract the contextual sequences and overall features of semantics, and residual connection was used to integrate features and further classify short texts. Experiments on two different datasets showed that ERNIE achieved a better classification effect than mainstream models, proving its feasibility and effectiveness.

1. Introduction

With the development of the mobile Internet in recent years, a large number of text data are generated every day from social apps such as instant messaging, search engines, and portal forums. And the number of short text data is on the rise. So, analyzing and processing such data have concerned scholars. Short text classification, a division of text classification, is a typical problem in natural language processing and machine learning. The purpose is to classify the content of short texts based on their labels. This classification method is widely used in intelligent question answering systems, emotional classification, commodity review analysis, spamming filtering, etc.

A piece of Chinese short text is as big as dozens of bytes and is characterized by sparsity, real-time response, and unstandardized use of language, due to which the effect of existing classification models is undesirable. Therefore, many researches have been done to improve the effect of Chinese short text classification. Traditional ways of extracting short text features were based on statistical approaches such as information gain, chi-square statistic, mutual information, and TF-IDF while combining machine learning to classify short texts. Li et al. [1] proposed a TFIDF method based on information gain and information entropy and studied feature

words in the same category and in different categories. The highlight of their study was that the distribution of feature words was one of the factors that influence the weight. In terms of chi-square statistic, Gao et al. [2] introduced a category-based CHI feature selection method by introducing distribution factors in the same category and in different categories. They applied this method to process low-frequency words in order to reduce their weight, with attention paid to the distribution of a feature word in different categories. In terms of mutual information, Wang and Qiu [3] proposed a hybrid feature selection algorithm by introducing frequency factors and regulatory parameters of feature words to improve the chi-square statistic and mutual information methods. These traditional feature representation methods involve many feature dimensions, which can easily lead to the “curse of dimensionality”. Moreover, as these methods hardly dealt with context semantics, they failed to achieve a desirable result in the effect of short text classification.

Mikolov et al. [4] proposed Word2Vec in 2013, a model based on word embedding. This model converted the text into a fixed-length vector, which could fundamentally solve the “curse of dimensionality” emerged in traditional feature representation methods while retaining the semantic relation between words [5]. Text vectors obtained through Word2Vec training showed a better classification effect than

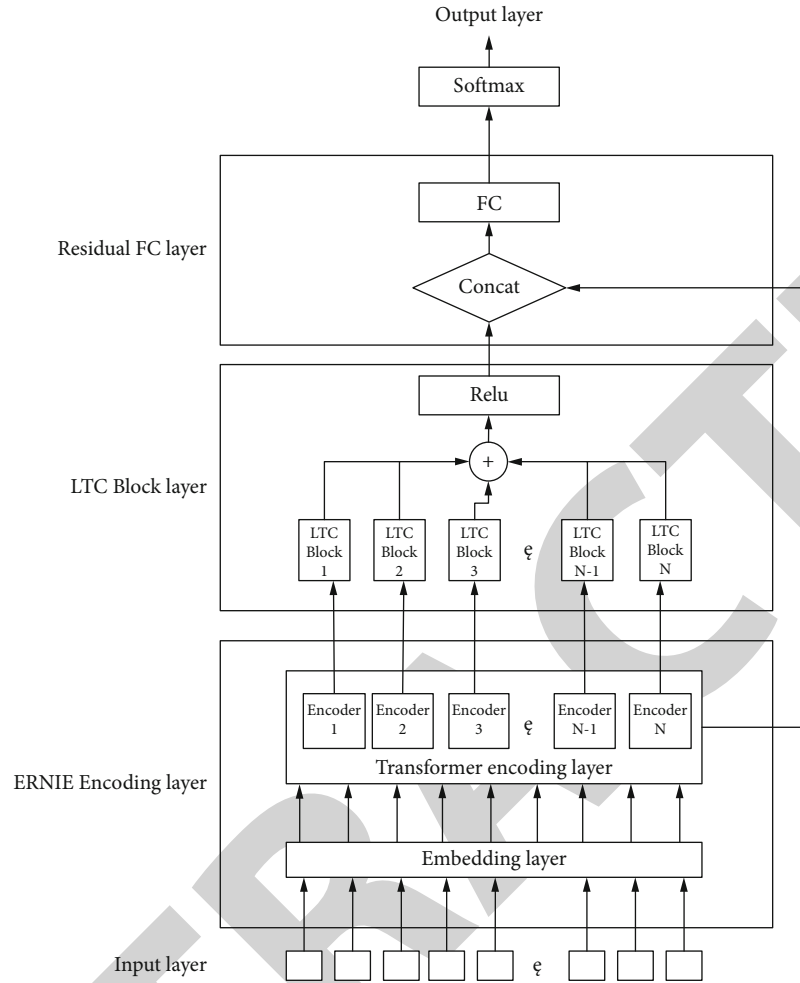


FIGURE 1: Structure of ERNIE based on LTC_Block.

traditional classifiers. Fesseha et al. [6] proposed a text classification method based on convolutional neural network (CNN) and low-resource language vocabulary embedding and constructed the CNN of continuous bag of words (CBOW) and the CNN of skip-gram. This method achieved a better effect than those using traditional machine learning methods. Gao et al. [7] extracted the short text features of different granularity with an improved CNN. In this way, key information with high classification accuracy could be effectively extracted. Although CNN and its methods could effectively extract part of the key information and yielded a good classification effect, context semantics [8] was still neglected. When processing natural language, long-short-term memory (LSTM) [9] is usually used to extract the context semantics of texts. Therefore, Ibrahim et al. [10] proposed a hybrid neural network in which they extracted text features and contexts of different granularity using CNN and BiLSTM and then integrated the output of the two parts. This algorithm greatly improved the effect of model prediction since it leveraged the advantages of both CNN and LSTM in text feature extraction.

After 2018, two pretrained models, BERT and ERNIE, were proposed, making major breakthroughs in multiple natural language processing tasks [11, 12]. To address the

polysemy problem in traditional language models, Shen and Ju [13] used BERT to extract the semantic features of microblog comments and then input them into BiLSTM for microblog tendency classification. BERT was proved effective in extracting semantic features. Qi et al. [14] proposed a method integrating TextCNN with ERNIE, which obtained text word embedding vectors using ERNIE and extracted features through convolution kernels of different sizes. It presented a better classification effect than Word2-Vec and the algorithm with BERT as the word embedding vector encoder. Lei and Qian [15] proposed an innovative BiGRU layer added to the ERNIE model. While the context information of the word is retained, adjustment can be made according to the ambiguity of the word, thus enhancing the semantic representation of the word. The structure was proved effective for text classification.

In summary, the main contributions of this paper are as follows:

- (1) ERNIE trained with prior knowledge was selected as the feature vector generation model to address the problem of polysemy which traditional embedding models failed to solve

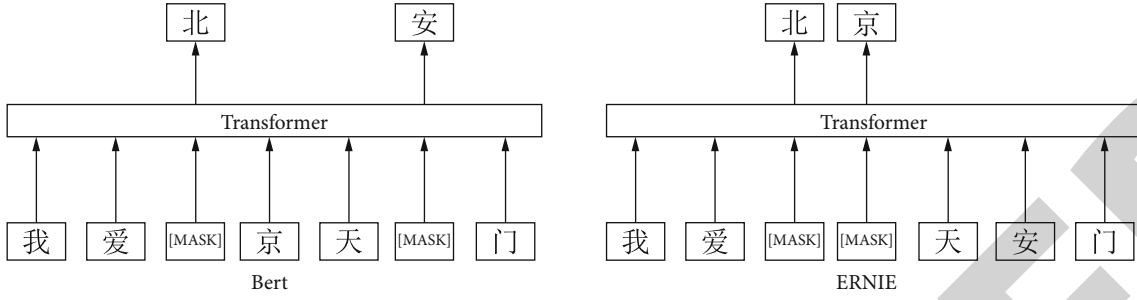


FIGURE 2: Difference between ERNIE and BERT in masking the contents.

- (2) The LTC_Block structure unit was proposed, combining the advantages of BiLSTM in extracting sequence context feature and those of TextCNN in extracting text local feature to form a two-channel structure
- (3) The LTC_Block was used to construct a layer of multiblock structure; under the ERNIE model, features were extracted at each layer of transformer, so as to solve the problem of losing feature information when data is processed through each transformer

After tested on different datasets, it is obvious that the feature vectors obtained from the proposed model in this paper are more extensive and have higher quality, with notably higher accuracy compared with mainstream text classification models.

2. LTC_Block-Based ERNIE

To enhance the capacity of ERNIE in extracting features and improve the effect of short text classification, this paper used LTC_Block, a double-channel structural unit composed of BiLSTM and TextCNN, to further extract features based on the feature coding of ERNIE. As is shown in Figure 1, ERNIE consists of three parts. Part 1 is the ERNIE encoding layer. The text corpus is input to preprocess and generate word vectors through the embedding layer, and eigenvectors are obtained through multiple transformer blocks. Part 2 is the LTC_Block layer. The output of transformer block is taken as the input to extract features from multiple dimensions. And the output of each LTC_Block is merged. Part 3 is the residual FC layer. The eigenvector output by LTC_Block is put together with the vector at the position of [CLS] in the output of ERNIE encoding layer through a way similar to residual connection. The matrix is output through the fully connected layer (FC). And the classification results are obtained through Softmax.

2.1. ERNIE Encoding Layer. Similar to BERT, ERNIE (Enhanced Language Representation with Informative Entities) also used the multilayer transformer as the basic encoder. But ERNIE is different from BERT in how the contents are masked. BERT randomly masks the corpus of a word or character so that it is trained to infer the masked contents; while ERNIE masks the phrases composed of several words or characters based on enhanced prior knowl-

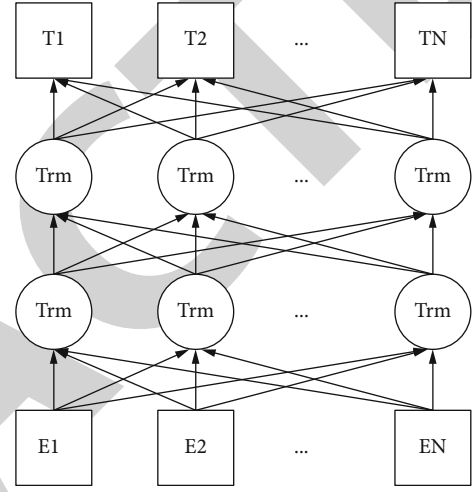


FIGURE 3: Structure of ERNIE.

edge, so it functions better in knowledge inference, as is shown in Figure 2.

ERNIE obtains context semantics [16] using a multihead self-attention composed of a number of self-attentions. The formulas are as follows:

$$\text{Head}_i = \text{Attention}\left(QW_i^q, KW_i^k, VW_i^v\right), \quad (1)$$

$$\text{MultiHead}(Q, K, V) = \text{Concat}_i(\text{Head}_i),$$

where i is the serial number of head; W_i^q , W_i^k , and W_i^v are training weight parameters; Q , K , and V are the three vector matrices of query, key, and value that can show the correlation between a single feature point and other feature points. The structure of ERNIE is shown in Figure 3. After the short text is preprocessed, the embedding layer is converted to a vector via transformer encoder that contains the multihead attention. Different versions of ERNIE use different numbers of heads. The base version of ERNIE consists of 12 heads [17].

The feature vectors calculated based on multihead attention indicate that feature words rely on context. Therefore, the feature vectors of the same feature word are calculated to a different result in different corpora. Such as:

Text1: “我喜欢吃苹果。”

Text2: “我喜欢苹果手机。”

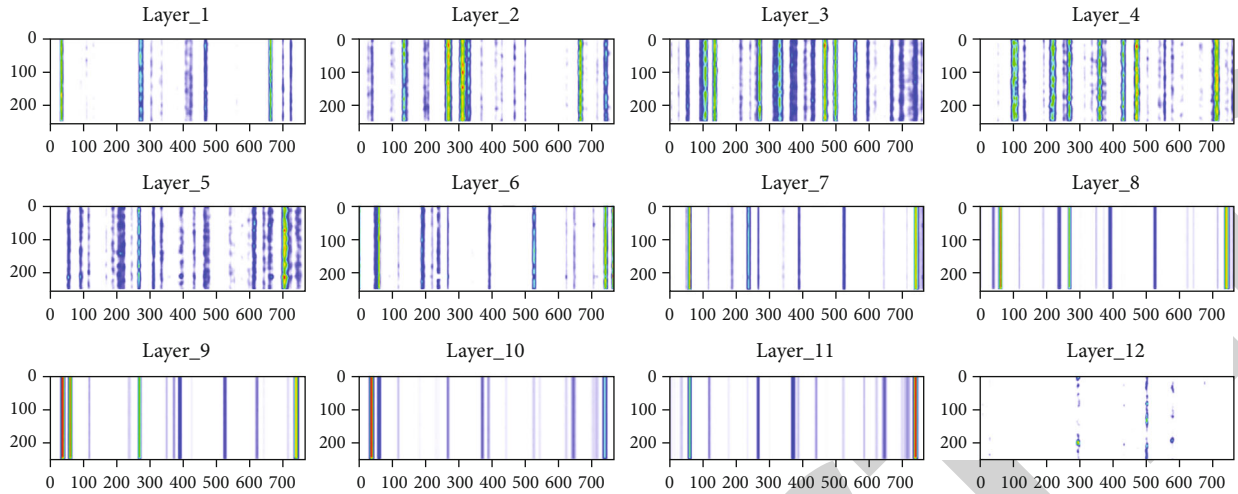


FIGURE 4: Heatmap of feature obtained from 12 transformer layers.

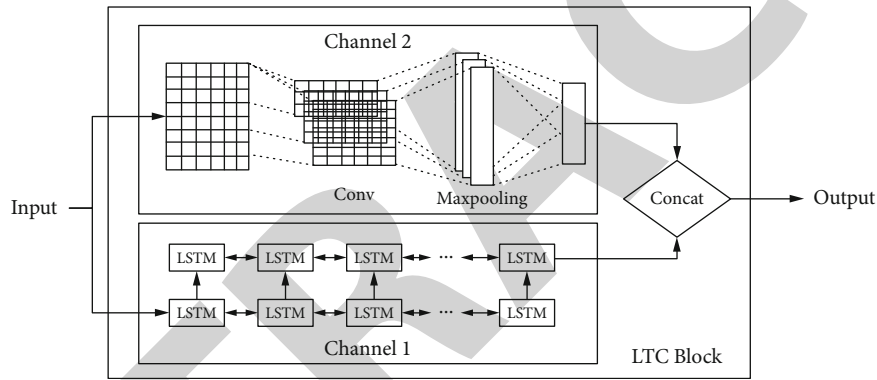


FIGURE 5: Structure of LTC_Block.

TABLE 1: Distribution of datasets.

Category	THUCNews			Category	Toutiao dataset		
	Sample number of training set	Sample number of validation set	Sample number of experiment set		Sample number of training set	Sample number of validation set	Sample number of experiment set
Finance	18000	1000	1000	Agriculture	15000	1000	1000
Life	18000	1000	1000	Auto	15000	1000	1000
Stocks	18000	1000	1000	Culture	15000	1000	1000
Education	18000	1000	1000	Education	15000	1000	1000
Science	18000	1000	1000	Entertainment	15000	1000	1000
Society	18000	1000	1000	Finance	15000	1000	1000
Politics	18000	1000	1000	Game	15000	1000	1000
Sports	18000	1000	1000	Housing	15000	1000	1000
Game	18000	1000	1000	Military	15000	1000	1000
Entertainment	18000	1000	1000	Sports	15000	1000	1000
				Tech	15000	1000	1000
				Travel	15000	1000	1000
				World	15000	1000	1000

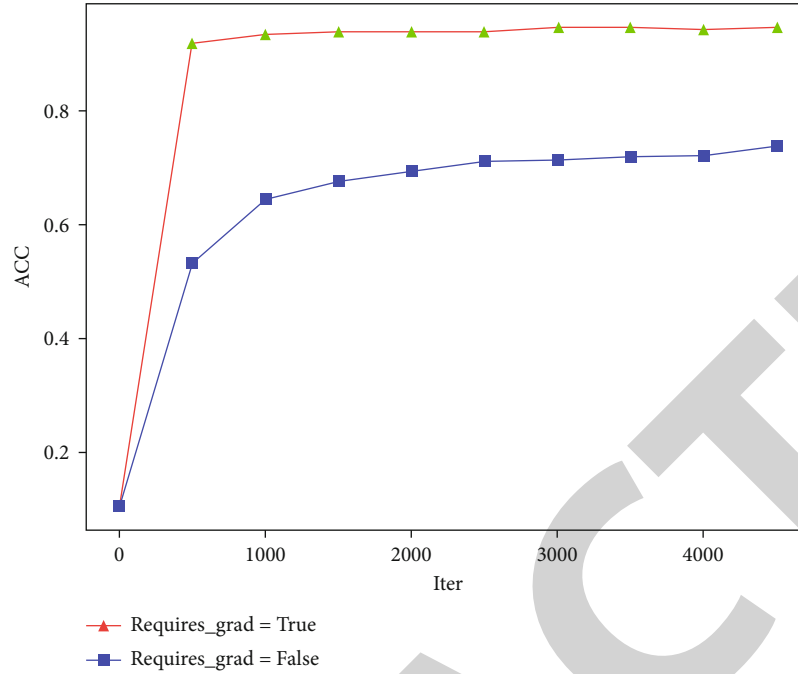


FIGURE 6: Impact of fine-tuning of parameters of the pretrained model on the classification effect.

TABLE 2: ERNIE encoding layer and training parameter settings.

Name of parameter	Value
Number of layer	12
Hidden size	768
Heads	12
Vocab size	18000
Hidden_act	Relu
Padding size	32
Learning_rate	5e-5
Batch size	256
Num_epoch	10

TABLE 3: Settings of double-channel parameters in the LTC_Block.

Channel 1		Channel 2	
Name of parameter	Value	Name of parameter	Value
Hidden_size	768	Filter_sizes	(3, 5, 7)
Rnn_hidden	256	Num_filters	256
Num_layers	2		

TABLE 4: Confusion matrix.

		Prediction result	
		Positive	Negative
Actual result	Positive	TP (true positive)	FN (false negative)
	Negative	FP (false positive)	TN (true negative)

In the two text corpora mentioned above, the same word “苹果”, in fact, represents different meanings. In text 1, it means fruit, while in text 2, it means a brand of smartphone.

The vectors obtained through the multihead attention mechanism are:

$$V_{\text{Text1}}[\text{“苹果”}] = [[0.88881, -0.88703, \dots, -0.28901], [7.4673e-04, -2.8027, \dots, 0.38405]].$$

$$V_{\text{Text2}}[\text{“苹果”}] = [[0.8563, -0.8001, \dots, -0.2527], [-0.0936, -1.3222, \dots, -0.2806]].$$

The feature word vectors were calculated to a different result, which can be used to address polysemy in practical calculation. However, the self-attention mechanism has its limitations. For example, the contextual sequences and overall features cannot be extracted. Moreover, under ERNIE model, some information will be missed in the process of extracting features from 12 transformer coding layers, thus reducing the information contained in the feature vector. Figure 4 shows the heatmap of feature obtained from each transformer layer. Therefore, in order to retain the feature information extracted from each transformer coding layer and expand the feature information of different dimensions, this paper proposes a structure based on the LSTM and TextCNN to make up for the deficiency of multihead attention.

2.2. LTC_Block Layer. LTC_Block is composed of two channels, as is shown in Figure 5. Channel 1 is BiLSTM (bidirectional long-short-term memory), which captures the hidden state of short text features at each moment through the combination of double hidden layers to capture the sequence relationship between features. Channel 2 is TextCNN (text convolutional neural networks), which extracts the overall corpus features with the convolution and pooling layers of different sizes. Finally, the outputs of the two channels are put together through CONCAT. The formula is as follows:

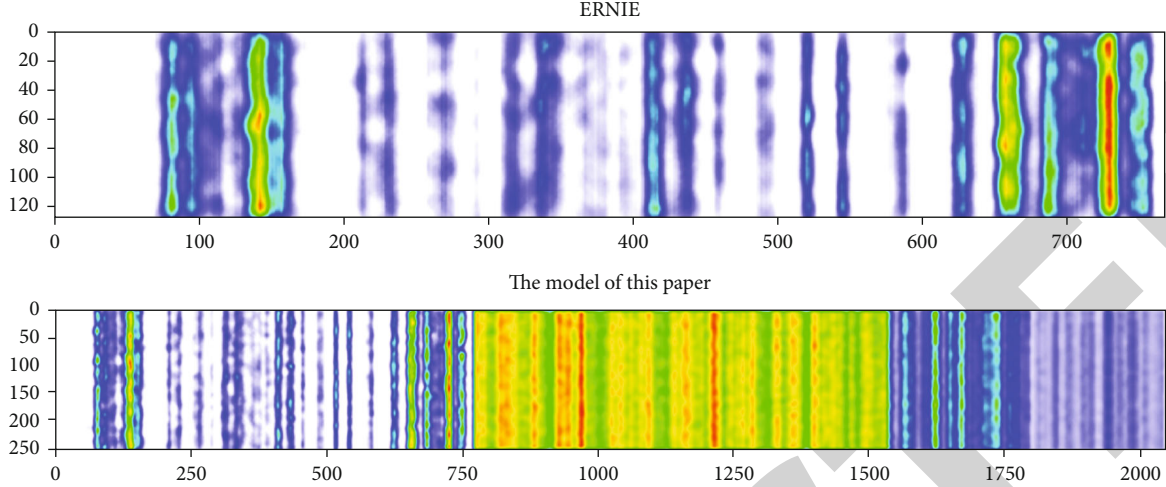


FIGURE 7: Heatmap of feature extracted from ERNIE layer and LTC_Block layer.

$$\text{LTC_Block_output}_i = \text{Channel1_BiLSTM}(T_i) \oplus \text{Channel2_TCNN}(T_i), \quad (2)$$

where T_i is the output of no. i transformer in the ERNIE encoding layer; \oplus is the combination of vectors. In the model, each transformer corresponds to one LTC_Block. All the LTC_Block_outputs obtained by Formula (2) are added up, and the output of the LTC_Block layer is obtained through Relu, an activation function. The formula is as follows:

$$\text{LTC_OUTPUT} = \text{Relu} \left(\sum_{i=0}^N \text{LTC_Block_output}_i \right). \quad (3)$$

3.2. Residual FC Layer. As is described in the previous two sections, the model proposed in this paper used multiple deep networks to extract features, so an attenuation of information would occur when the information was passed downwards. To minimize the loss of features caused by this attenuation, we proposed to use a method similar to residual connection in ResNet, a deep learning model, to put the information together. The eigenvector at the position of [CLS] output by ERNIE in 1.1 is put together with the LTC_OUTPUT eigenvector obtained in 1.2. The formula is as follows:

$$\text{Residual} = \text{Concat} \left(\text{ERNIE}_{[\text{CLS}]}, \text{LTC_OUTPUT} \right). \quad (4)$$

The residual is used as the eigenvector extracted by the model. After the eigenvector goes through FC and Softmax, the results of short text classification can be produced.

3. Experimental Results and Analysis

3.1. Introduction of Experimental Data. The experimental data of this paper was THUCNews published by the Natural Language Processing and Social Humanities Computing Laboratory of Tsinghua University and news dataset from Toutiao. The Toutiao dataset contained 382,688 pieces of data under 15 categories, and THUCNews contained more

than 60,000 pieces of data under 10 categories. The number of the piece of data under some categories of the Toutiao dataset was far less than that of other categories. In order to minimize the impact of uneven data distribution on model training, we extracted part of the data from each category of the Toutiao dataset. To ensure the consistency of the training set, experiment set, and validation set for each category, all the data were randomly extracted, and the data for the training set, experiment set, and validation set must be independent from each other. The distribution of datasets is shown in Table 1.

3.2. Experimental Environment and Parameter Setting. The experiments in this paper were carried out on Pyhton3.6. The model was written through pytorch1.7.1, a deep learning framework, and it was trained using Tesla P100, a graphics card.

This paper used the open Chinese base version of ERNIE, whose classification effect was not desirable due to the lack of prior knowledge for the current classification task. Figure 6 is the test with ERNIE+FC on THUCNews. And the test accuracy of the pretrained model with and without fine-tuning of parameters is 94.53% and 74.93%, respectively. Thus, fine-tuning parameters with prior knowledge significantly improves the classification effect of the current task. As for parameter configuration, when the pretrained model was loaded, the attribute value of requires_grad was set to be true, so that parameters could be fine-tuned in the training. The settings for other parameters are shown in Table 2.

The settings of double-channel parameters in the LTC_Block are shown in Table 3.

3.3. Experimental Results and Analysis

3.3.1. Evaluation Indicators. In this paper, three evaluation indicators, namely, P (precision), R (recall), and $F1$ were used, and their calculation formulas were, respectively, defined as $P = TP/(TP + FP)$, $R = TP/(TP + FN)$, $F1 = (2 \times P \times R)/(P + R)$. The confusion matrix is shown in Table 4.

TABLE 5: Comparison of the proposed model with the mainstream short text classification models in terms of experimental results by dataset.

Model	THUCNews			Toutiao dataset		
	<i>P</i>	<i>R</i>	<i>F1</i>	<i>P</i>	<i>RP</i>	<i>F1</i>
TextCNN	0.9075	0.9072	0.9071	0.8551	0.8538	0.8541
TextRCNN	0.9093	0.9090	0.9089	0.8408	0.8396	0.8394
DPCNN	0.9120	0.9111	0.9113	0.8338	0.8325	0.8325
TextCNN_Att	0.8943	0.8946	0.8944	0.8440	0.8431	0.8429
ERNIE+FC	0.9453	0.9451	0.9451	0.8941	0.8942	0.8940
Bert+FC	0.9461	0.9457	0.9458	0.8913	0.8914	0.8912
ERNIE+RCNN	0.9442	0.9441	0.9441	0.8944	0.8931	0.8930
The model of this paper	0.9504	0.9503	0.9503	0.9003	0.8995	0.8994

3.3.2. *Comparative Analysis of Experimental Results.* The feature vectors obtained from the proposed model in this paper are more extensive and have higher quality. As shown in Figure 7, the comparison of feature heatmap extracted from LTC_Block layer and ERNIE layer shows that the feature dimension obtained by the proposed model is nearly three times higher with more abundant feature content. It not only has the same content differentiation features of polysemy as in ERNIE model but also contains the overall information features of the text and semantic sequence.

To validate the classification performance, we compared the model with mainstream short text classification models in terms of *P* (precision), *R* (recall), and *F1*. The comparison results are shown in Table 5.

As is shown in Table 5, among the mainstream short text classification models, the pretrained models such as ERNIE or BERT achieve a good classification effect, while those traditional models such as TextCNN or RNN have a poor classification effect. This shows that the semantic-based feature extraction methods have an important impact on the classification effect. ERNIE, an LTC_Block-based short text classification model proposed in this paper, has expanded features extracted by the pretrained model. From experiment results from both datasets, we can conclude that the classification effect of ERNIE is better than mainstream models.

4. Conclusion

This paper proposed an LTC_Block-based short text classification model named ERNIE to classify Chinese short texts and extract semantics in the corpus to address the polysemy problem in the text. In this model, LTC_Block, a double-channel structural unit composed of BiLSTM and TextCNN, was used to extract the contextual sequences and overall features of semantics, and residual connection was used to integrate features and further classify short texts. The experiments on both datasets showed that the classification effect of ERNIE was better than mainstream short text classification models. According to this paper, ERNIE can be applied to other natural language processing tasks, such as long text classification and emotion analysis, to verify its effect in different application scenarios.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

It is declared by the authors that this article is free of conflict of interest.

References

- [1] X. M. Li, H. R. Li, L. Xue, and G. J. He, "TFIDF algorithm based on information gain and information entropy," *Computer Engineering*, vol. 38, no. 8, pp. 37–40, 2012.
- [2] G. Baolin, Z. Zhiguo, Y. Wenwei, and X. Zeli, "Feature selection method based on combination of category and improved CHI," *Application Research of Computers*, vol. 35, no. 6, pp. 1660–1662, 2018.
- [3] Z. Wang and X. H. Qiu, "An improved text feature selection method mixed CHI with MI," *Computer Technology and Development*, vol. 28, no. 4, pp. 87–90–87–94, 2018.
- [4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in Neural Information Processing Systems*, vol. 26, pp. 3111–3119, 2013.
- [5] X. Y. Niu and E. Y. Zhao, "Research on Chinese weibo text classification based on Word2Vec," *Computer Systems & Applications*, vol. 8, pp. 256–261, 2019.
- [6] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, "Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya," *Information*, vol. 12, no. 2, p. 52, 2021.
- [7] Y. L. Gao, C. Wu, and M. Zhu, "Short text classification model based on improved convolutional neural network," *Journal of Jilin University*, vol. 58, no. 4, pp. 923–930, 2020.
- [8] A. C. Metlapalli, T. Muthusamy, and B. P. Battula, "Classification of social media text spam using VAE-CNN and LSTM model," *Ingénierie des Systèmes D Information*, vol. 25, no. 6, pp. 747–753, 2020.
- [9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] M. A. Ibrahim, M. U. G. Khan, F. Mehmood, M. N. Asim, and W. Mahmood, "GHS-NET a generic hybridized shallow neural

- network for multi-label biomedical text classification,” *Journal of Biomedical Informatics*, vol. 116, no. 1, article 103699, 2021.
- [11] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” 2019, <https://arxiv.org/abs/1810.04805>.
- [12] Y. Sun, S. Wang, Y. Li et al., “ERNIE: enhanced representation through knowledge integration,” 2019, <https://arxiv.org/abs/1904.09223>.
- [13] Z. Q. Shen and T. Ju, “Research on tendency analysis of micro-blog comments based on BERT and BLSTM,” *Information Studies: Theory & Application*, vol. 43, no. 8, pp. 173–177, 2020.
- [14] J. Q. Qi, C. Y. Chi, and X. G. Zhan, “ERNIE-CNN text classification model,” *Journal of University of Science and Technology Liaoning*, vol. 44, no. 1, pp. 56–61, 2021.
- [15] J. H. Lei and Y. Qian, “Chinese-text classification method based on ERNIE-BiGRU,” *Journal of Shanghai University of Electric Power*, vol. 36, no. 4, pp. 329–335–329–350, 2020.
- [16] A. Vaswani, N. Shazeer, N. Parmar et al., “Attention is all you need,” in *NIPS2017*, pp. 6–12, Long Beach, CA, USA, 2017.
- [17] W. Zheng, X. Liu, and L. Yin, “Sentence representation method based on multi-layer semantic network,” *Applied sciences*, vol. 11, no. 3, p. 1316, 2021.