

Article

Research on Long Text Classification Model Based on Multi-Feature Weighted Fusion

Xi Yue ^{1,2}, Tao Zhou ¹, Lei He ^{1,2,*} and Yuxia Li ^{3,*}¹ School of Software Engineering, Chengdu University of Information Technology, Chengdu 610225, China; yuexi@cuit.edu.cn (X.Y.); 3200706015@stu.cuit.edu.cn (T.Z.)² Sichuan Province Engineering Technology Research Center of Support Software of Informatization Application, Chengdu 610225, China³ School of Software Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

* Correspondence: helei1978@cuit.edu.cn (L.H.); liyuxia@uestc.edu.cn (Y.L.)

Abstract: Text classification in the long-text domain has become a development challenge due to the significant increase in text data, complexity enhancement, and feature extraction of long texts in various domains of the Internet. A long text classification model based on multi-feature weighted fusion is proposed for the problems of contextual semantic relations, long-distance global relations, and multi-sense words in long text classification tasks. The BERT model is used to obtain feature representations containing global semantic and contextual feature information of text, convolutional neural networks to obtain features at different levels and combine attention mechanisms to obtain weighted local features, fuse global contextual features with weighted local features, and obtain classification results by equal-length convolutional pooling. The experimental results show that the proposed model outperforms other models in terms of accuracy, precision, recall, F1 value, etc., under the same data set conditions compared with traditional deep learning classification models, and it can be seen that the model has more obvious advantages in long text classification.



Citation: Yue, X.; Zhou, T.; He, L.; Li, Y. Research on Long Text Classification Model Based on Multi-Feature Weighted Fusion. *Appl. Sci.* **2022**, *12*, 6556. <https://doi.org/10.3390/app12136556>

Academic Editor: Federico Divina

Received: 7 June 2022

Accepted: 27 June 2022

Published: 28 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: text classification; multi-feature fusion; pretraining model; neural networks

1. Introduction

In today's exponential growth of data, the processing and summarization of massive text data on the Web have become an important issue, and how to organize and classify the content diversity of different data in various fields has been a common concern. Text classification is an important research direction in the field of natural language processing. Text categorization is based on the feature extraction and analysis of text data and is classified automatically according to the preset classification system. Text tasks [1–5] are of great importance in several fields. For the text task of Chinese, due to the intricate semantic information and strong correlation of contextual information, and according to the length of the text, it can be divided into long text and short text. The semantic complexity of the long text and the difficulty of extracting relevant features also affect the effect of text classification.

Researchers have now proposed various classification algorithms and models based on machine learning and deep learning and achieved good classification results. With the development of machine learning, researchers have replaced the traditional manual classification method with machine learning. Applied it to text classification tasks by improving machine learning algorithms [6–10].

With the introduction of the development of deep learning [11–14] into the field of text classification, the research on classification methods has achieved great success. Kim [15] proposed to use the trained word vectors for feature extraction using convolutional neural networks. However, the convolutional neural networks can only extract local feature

information and cannot complete full-text features. R Johnson et al. [16] proposed a deep convolutional neural network combined with residual connectivity [17] to alleviate the DPCNN model of deep networks, which solved the problem of feature dependence on the long-distance text and effectively improved the semantic depth of word features. Pota et al. [18] proposed to analyze the learning process of convolutional networks used for problem-solving classification and verified that the optimization resulted in higher classification accuracy.

Based on long short-term memory network [19] can add location information to obtain full-text semantic structure. Chang Xu [20] et al., proposed an improved variant multi-channel RNN model for dynamically capturing and exploiting local semantic structure information and combined with an attention mechanism to capture local structure and dependencies in sentences. Du Jie et al. [21] proposed to learn multiple information simultaneously through a gating mechanism with LSTM, such as sequence information and lexicality, to achieve high accuracy in classification.

Some scholars proposed to introduce the attention mechanism applied in the image domain into text classification to evaluate the importance of each feature to the current task and distinguish the secondary features from the primary features. Zhou et al. [22] proposed an attention-based mechanism combined with BiLSTM to capture important semantic information in sentences; it can improve the efficiency of the text representation task. Zheng et al. [23–25] proposed a multi-headed attention-based mechanism-based semantic representation network, which can obtain semantic information at different levels to enhance the semantic representation of sentences.

To better feature extraction of text, the proposed pre-training model has become the main research approach for text-related tasks. The BERT pre-training model proposed in 2018 [26] is a bidirectional deep language model based on the Transformer model [27]. Its attention mechanism can capture bidirectional contextual semantics, and BERT processes word vectors through a pre-training model, solving the problem of polysemy of static word vectors. Subsequent researchers have improved the BERT model [28,29], but all of them have problems, such as requiring more training data and training time, or the performance of the model also decreases.

To address the problems of incomplete feature representation and difficulty in capturing text dependencies over long distances in long text classification tasks. Sanda et al. [30] conducted a comparative analysis of text feature representation models and verified the impact of text feature representation on the classification effect. In addition, multiple network models were combined [31–34] to apply feature fusion to text classification tasks with good results. Jang et al. [35] proposed a hybrid model of BiLSTM and CNN based on the attention mechanism, which captures the correlation of adjacent words to improve the classification accuracy. Abdi et al. [36] proposed a multi-feature fusion approach based on LSTM for classification, which overcomes the problems such as word order and information disappearance in traditional methods.

In this paper, on studying various current feature vector processing and classification models, we summarize and analyze the advantages of feature vector representation for text tasks and propose a new study of text classification based on multi-feature weighted fusion extraction for feature extraction and classification effect of long text. The global features containing contextual information are obtained by processing the text data with the BERT model, and the convolutional pooled features are fused with the features from the upper-level convolution in the TextCNN model, and the weighted representation is combined with the attention mechanism. The global features and multi-level weight local features are stitched together and then input into the equal-length convolutional block of the convolutional pooling layer of the network structure for feature extraction, which solves the text long-distance dependency problem and uses the obtained feature vectors as classification inputs, and finally obtains the classification results. The results show that all the metrics are better than the traditional deep learning classification models under the same dataset conditions.

2. Materials and Methods

2.1. DataSet

The experimental data in this paper use the Chinese text classification THUCNews [37] open-source dataset introduced by the Natural Language Processing Laboratory of Tsinghua University and the Sohu all-news dataset [38] compiled by Sogou Lab.

THUCNews was generated by filtering and filtering historical data from the Sina News RSS feeds between 2005 and 2011 and contains 740,000 news documents (2.19 GB), all in UTF-8 plain text format. Based on the original Sina News classification system, we re-integrated and divided 14 candidate classification categories: finance and economics, lottery, real estate, stocks, home, education, technology, society, fashion, current affairs, sports, horoscope, games, and entertainment.

The data set compiled by Sogou Lab comes from news data from 18 channels, including domestic, international, sports, social, entertainment, finance, education, science, and technology, during 2012 in Sohu News, and is organized according to the data format to provide URL and body information, for example Table 1.

Table 1. Sample data of SoHu News.

Data Format
<pre> <doc> <url>Page URL</url> <docno>Page ID</docno> <contenttitle>Page title</contenttitle> <content>Page content</content> </doc> </pre>

In this paper, the ten news categories of finance, real estate, education, technology, society, current affairs, sports, horoscope, games, and entertainment, which have the largest audience in the two datasets, are selected respectively.

2.2. Methods

In this paper, we propose a long text classification model based on multi-feature weighted fusion, like Figure 1, in which the word embedding representation of the model is vectorized by the word embedding layer of the BERT model, and then the word vector representation is passed through the parallel network structure of BERT and convolutional neural network, respectively, to extract global semantic features according to the efficient learning of text semantics by the BERT model, and the local features of the text sequence by the TextCNN model. The global features and semantic associations of text are obtained and the weighted local features are obtained at the same time, and the global features and the weighted local features are vector fused and obtain the multi-feature word vector representation. Compared with single features, the multi-feature representation can obtain global contextual semantic and local important information and reduce the secondary feature representation by weighting the local features, which improves the diversity of feature representation and enhances the acquisition of text information. The vector representation of multi-features is input to capture long-distance-dependent equal-length convolutional blocks for feature extraction, and finally, the text classification results are obtained by fully connected layers combined with Softmax.

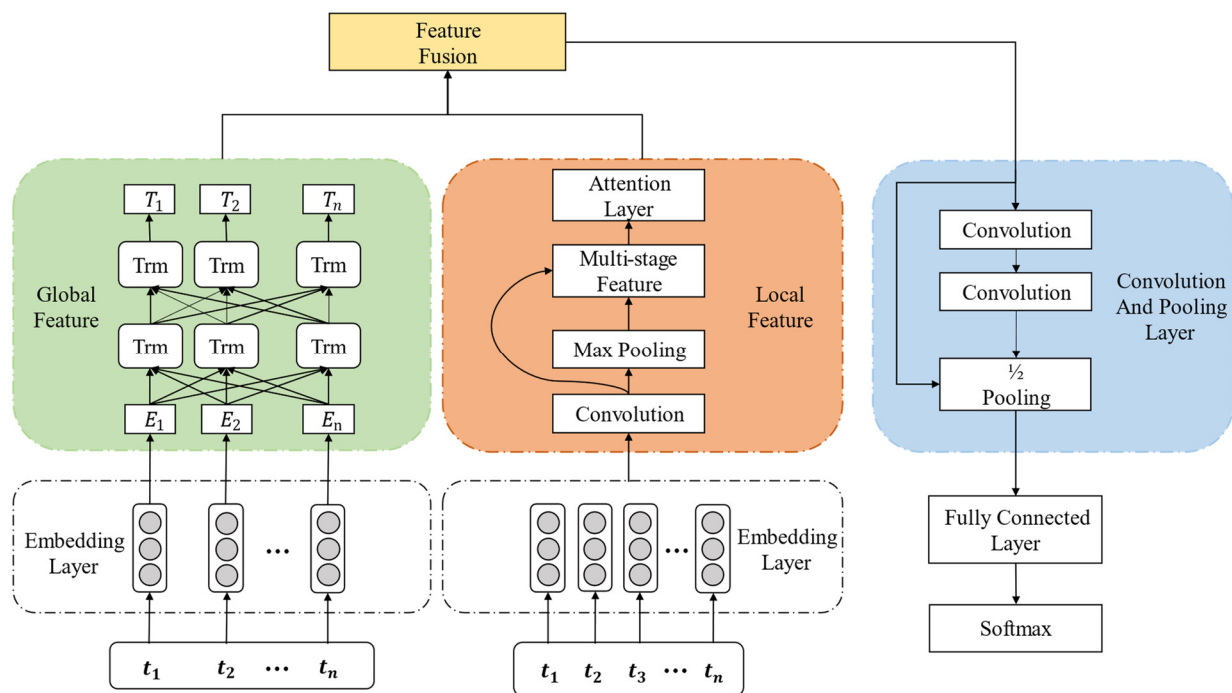


Figure 1. BTCDC network structure.

2.2.1. Word Embedding Layer

Word embeddings are intermediate representations that capture the semantic patterns between words, and the paper [39] verified that more informative word embedding representations could effectively improve the accuracy of classification tasks. The input vector of the BERT model is stitched by three vectors: Word Embedding, Segment Embedding, and Position Embedding, and [CLS] and [SEP] are added at the head and tail of the text, it indicates the beginning and end position of text information respectively. Word embedding transforms words in the text into vectors of fixed dimensions. Two special flags are inserted at the beginning and end. Sentence segment embedding can deal with sentence pairs of input text, representing the semantic similarity of multiple sentences, and different sentences have different feature representations. Position embedding encodes the position information of words in the text into a vector. Unlike Transformer, which uses a fixed sine and cosine function, the position embedding in BERT is obtained by model learning. The word embedding representation of the BERT model is obtained by adding the three vector embedding representations.

2.2.2. Feature Fusion Layer

TextCNN is a representative neural network model that introduces a convolutional neural network into natural language processing for text classification. The model combines the trained language model with a neural network, captures the local features of text vectors by the convolutional neural network, and automatically combines and filters the features by convolutional kernels of different sizes to obtain different text feature information. Through the pooling layer, the main feature information is kept unchanged while the dimension is further reduced, which effectively prevents the overfitting of the model and improves the generalization ability of the model. In the pooling selection, the maximum pooling is used to extract the most effective feature information in the text.

The input word vector is represented in its model as:

$$V_{(i:n)} = v_1 \oplus v_2 \oplus \cdots \oplus v_n \quad (1)$$

where v_i denotes the k -dimensional representation of the i th word in the text in the vector, \oplus denotes the word vector connection, n denotes the number of words in the text, and the

stitched text vector-matrix $V_{i:n}$ is fed into the convolution layer to extract local features through the convolution kernel. The convolution process of text features is performed in one direction of the text sequence, and a word in the text has a certain word dimension when vector representation is performed, so the text sequence is an $n \times k$ matrix when vector representation is performed. The TextCNN network model contains several convolution kernels of different sizes during convolution, and the convolution formula is as follows.

$$c_i = \tanh(w \cdot v_{i:i+h-1} + b) \quad (2)$$

$$C = [c_1, c_2, \dots, c_{n-h+1}] \quad (3)$$

where w is the convolution kernel parameter, h is the convolution kernel width, $w \in R^{hk}$, b is the bias term, \tanh is the activation function, and C is the feature output after the convolution layer extraction.

Each convolution operation is equivalent to a feature vector extraction. By defining different convolution windows, different feature vectors can be extracted. After Max Pooling, the feature vectors are filtered for maximum features and stitched to form a text vector representation. The word vector matrix with local features is obtained by reducing the feature dimension of the pooling layer to select the maximum features in the region. While acquiring text features through convolutional pooling, to acquire more features at different levels, the features T_b obtained from the pooling layer are combined with the features T_a obtained from the higher convolutional layer, and obtain the multi-level features T in the text convolution process, and the information fusion of the two input channels is obtained by element-level summation, and the computational effort is simplified. Where T_i denotes the word vector dimension representation of the i -th word of the text.

$$T = [T_{a_1} + T_{b_1} \dots T_{a_n} + T_{b_n}] \quad (4)$$

where T_a is the features obtained by the first layer of convolution, T_b is the features obtained after pooling, and T is the multi-level feature output vector representation combined.

Weights are added according to the degree of local features' representation of text semantics in the attention mechanism, and the weight ratio of minor features in the multi-feature representation is reduced in the subsequent splicing with global features so that the influence of the secondary features on the spliced semantic feature representation is avoided. The weighted local features are obtained by attention calculation of the combined multi-level feature representation T through the attention mechanism. The initial weights are given to the vectors by attention and the weights are updated by training.

$$\beta_i = \tanh(WT_i + b) \quad (5)$$

Normalized probability representation of attention weights by Softmax function:

$$h_i = \frac{\exp(\beta_i)}{\sum_1^n \exp(\beta_k)} \quad (6)$$

The weighted local features H are obtained by attentional computation of the combined multilevel feature representation T through the attentional mechanism:

$$H = \sum_1^n h_i T_i \quad (7)$$

Weights are added according to the degree of local features' representation of text semantics in the attention mechanism, and the weight ratio of secondary features in the multi-feature representation is reduced in the subsequent splicing with global features, can avoid their influence on the semantic feature representation after fusion.

In the process of BERT pre-training, a large amount of dataset and computational resources are required. In this paper, we use the Chinese pre-training model Bert-base-Chinese provided by Google as the base pre-training model and fine-tune it on this basis.

The word vector matrix is passed through the Self-attention layer by the BERT model, which enables the Encoder to learn the contextual information of the text while encoding. Then it passes through a layer of Add&Norm layer, which is normalized to make the output with mean and standard deviation. After that, it goes to the fully connected layer of the feedforward neural network. Where x_i denotes the word vector representation of the i -th word.

$$X = (x_1, x_2, \dots, x_n) \quad (8)$$

Reference DenseNet [40] uses a dense concatenation mechanism to stitch the text feature vectors output from the two models, and two or more features are spliced according to the channel or dimension. By adding weights to the local feature vectors, it can prevent certain local features from affecting the representation of global features on text semantics, and finally get a new feature representation is finally obtained. In terms of feature information fusion, in this paper, while fusing global features and local features, the local features are weighted and modified in combination with the attention mechanism, as shown in Equation (9), with n being the vector dimension.

$$E(e_1, \dots, e_n) = [X_{x_1} \cdot H_{h_1}, \dots, X_{x_n} \cdot H_{h_n}] \quad (9)$$

The weight factor makes the features important to the task more obvious in the local features, reduces the influence of the secondary information in the local features on the global semantic information, and finally obtains the text semantic feature vector after multi-feature fusion.

2.2.3. Convolution Pooling Layer

When convolution is performed in conventional convolutional networks, often the text sequence length is continuously reduced and some long-distance text feature dependencies are lost. To solve this problem, the text sequence length is kept constant after extraction by equal-length convolution, and each word in the sequence and its left and right features are compressed and fused. Therefore, each word of the input sequence gets the contextual information of the words within a certain range to its left and right, and after convolution, each word also obtains a more accurate semantics reinforced by the contextual semantics of the range. After continuous deep convolution, the semantics of longer distance is captured and deeper semantic information is obtained.

In this paper, the equal-length convolutional block is proposed as feature extraction, and for the problems of long text with difficult information extraction and features that have been lost, its equal-length convolution is constituted as 2×1 convolutional pooling. Like Figure 2, the text sequence is subjected to equal-length convolution with $1/2$ pooling layer for feature extraction of the sequence, and the sequence output by the pooling layer is expanded twice for each word to capture the contextual semantics within a certain range. The convolutional block consists of two layers of equal-length convolution to improve the richness of the word embedding representation. After each convolutional block, a pooling layer of size 3 and step 2 is used to obtain the intermediate feature representation, and the cited ResNet [17] residual connection is introduced to prevent the gradient dispersion and the low weight of the convolutional layer by deepening the number of network layers.

The feature vector representation matrix E after multi-feature fusion is passed through equal-length convolution combined with $1/2$ pooling for feature extraction, and its network structure extracts long-distance dependencies of text and more feature information, and finally, the text feature vector is passed through a fully connected layer combined with Softmax to obtain the final classification results.

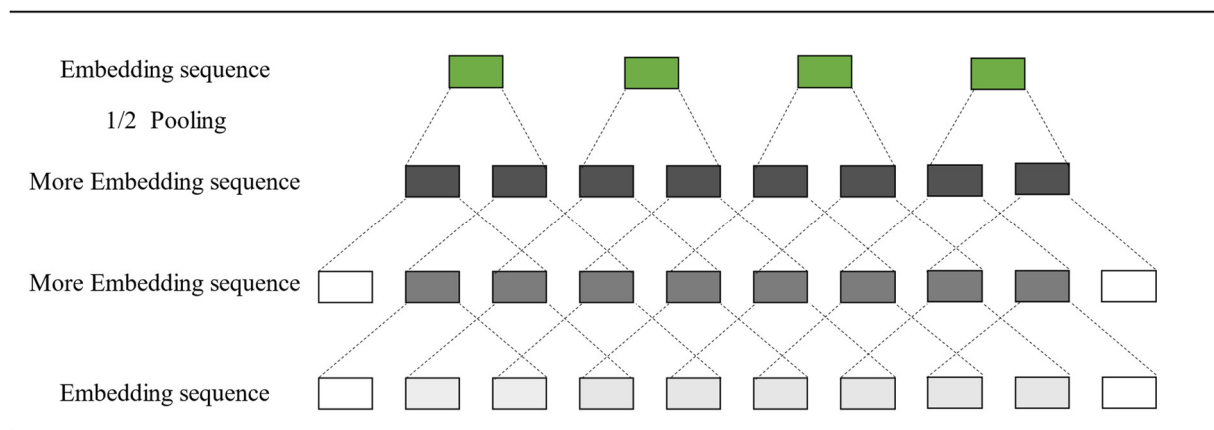


Figure 2. BTCDC Convolution Pooling Layer.

3. Results

3.1. Experimental Environment and Design

The experimental environment used in this paper is set up with an interpreter of Python 3.9, an operating system of Windows 10, a deep learning framework of Pytorch, a server CPU of Intel(R) Core(TM) i9-10900, a running memory of 32 G, and a GPU graphics card of NVIDIA GeForce GTX 2060. In the text classification, Adam optimizer was chosen for the task experiments [41]. To verify the classification effectiveness of the model on the classification task with other network models, several comparison models are introduced in the experiments in this paper, namely TextCNN [15], BiLSTM [19], BERT [26], DPCNN [16], BiLSTM+CNN [31], and BTCDC.

3.2. Scoring Criteria

The evaluation metrics of this experiment for the THUCNews open-source dataset are accuracy and Macro-averaging F1 scores for multiple categories. The macro-averaging calculates the F1 score for each category and then obtains the arithmetic mean of F1 values for all categories. The evaluation metrics are calculated as follows:

1. Accuracy

The accuracy rate in the classification task indicates the number of correctly classified samples as a percentage of the total number of samples:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (10)$$

2. Precision

The precision rate in the classification task indicates the number of correctly classified positive samples as a percentage of the number of samples judged to be positive:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

3. Recall

Recall in a classification task represents the number of correctly classified positive samples as a percentage of the true number of positive samples:

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

4. F1 value

The F1 value in the classification task indicates the performance of the combined evaluation accuracy and recall:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (13)$$

5. Macro Average

Macro averages in multiclassification tasks represent the average of indicators across all categories:

$$Macro_F = \frac{\sum_{i=1}^n F_i}{n} \quad (14)$$

In the above calculation equation. *TP* is the number of positive samples predicted to be positive, *FN* is the number of positive samples predicted to be negative, *FP* is the number of negative samples predicted to be positive, and *TN* is the number of negative samples predicted to be negative.

3.3. Experimental Results

3.3.1. Experimental Results on the THUCNews Dataset

All the above comparison models were experimented with and analyzed on the THUCNews open source dataset of Tsinghua University. To reduce the model errors, the results were compared for multiple experiments and the mean value of the total experimental results was taken for the experimental comparison. The accuracy and macro-average values of each model were obtained from the classification performance of each model on the dataset, and the experimental results of multiple models were compared in groups to verify the influence of different models on the classification results. After the experiment the experimental results of each model are shown in Table 2, and the experimental cases are shown in Table 3.

Table 2. Experimental results on the THUCNews dataset.

Model	Accuracy (%)	Macro-F (%)
TextCNN	93.25	93.17
BiLSTM	93.34	93.36
BERT	94.58	94.59
DPCNN	93.56	93.57
BiLSTM+CNN	93.73	93.71
BTDCD	95.27	95.26

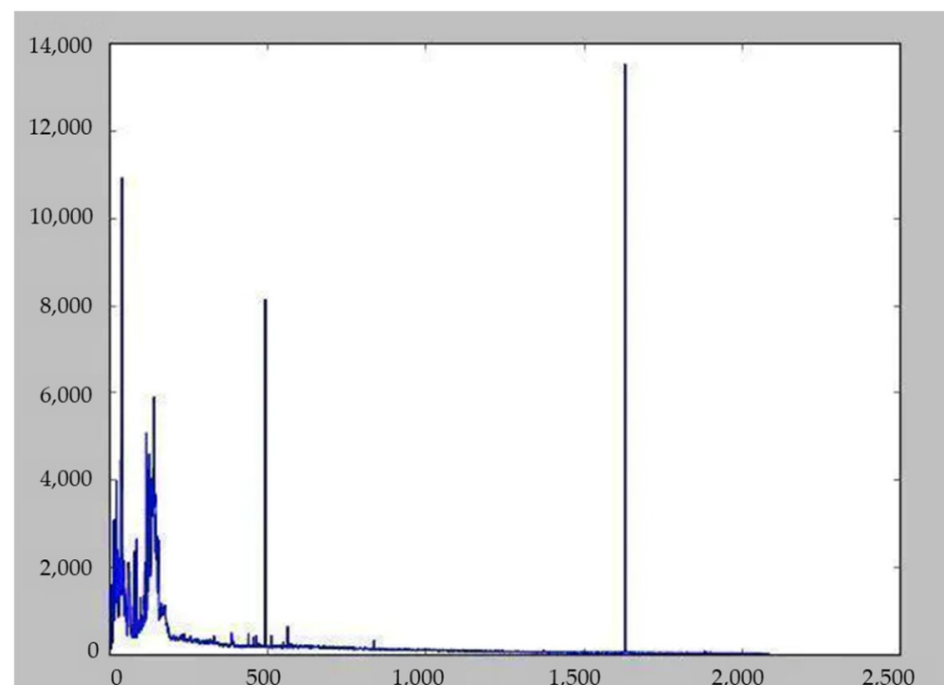
3.3.2. Experimental Results on Sohu News Dataset

To evaluate the effectiveness of the model results, this paper collates a sample of 100,000 Sohu all-news datasets collected by the Sogou Lab for secondary comparison experiments to verify the classification effect of the model under long text categories. If the dataset categories are the same as the above experimental categories, the length of the statistical data set text sequences is shown in Figure 3.

To avoid some interference factors caused by the imbalance of data amount, the data of different text lengths in the data set are counted. When sorting out the data set, the proportion of the number of short texts and long texts is modified, and the data set with the text sequence length of [500, 1600] is selected for experiments to verify the classification performance of the model under different length text data sets. The results of Experiment 2 are shown in the following table.

Table 3. Classification results for the THUCNews dataset (Chinese Expression).

Classification Result	Classification Result
<p>Text:</p> <p>意甲第36轮,国米主场4-2逆转恩波利,劳塔罗又进了2个球,其中本队第二个球扳平比分,随后第64分钟,他将比分改写为3-2,为国米拿下比赛立下汗马功劳。现在的劳塔罗状态太好了,最近11场比赛打入了11球,在意甲联赛中,他打入了19球,这是新高,虽然排在意甲第三,但是他罚点球少,只有3个,因莫比莱 (27球) 和弗拉霍维奇 (23球) 点球都多,前面进了7个,后面进了5个。所以运动战进球,劳塔罗进步很大。劳塔罗的爆发对于阿根廷世界杯绝对利好,有这样一个生猛的,梅西世界杯之旅稳当了吧? 梅西又多了个帮手是毋庸置疑的。劳塔罗,迪玛利亚、梅西绝对是顶配了。此外还应该有尤文的迪巴拉、佛罗伦萨前锋尼古拉斯冈萨雷斯、国际米兰的华金科雷亚、马竞的安赫尔科雷亚、塞维利亚边锋奥坎波斯、当然还有大巴黎的伊卡尔迪,听着都是那么让人羡慕。所以本届世界杯之旅,阿根廷的锋线绝对可以和任何一支球队比进球,当然能拿下冠军还要看防守,不过能不能比对方多进一球,那真要看梅西带领的锋线如何表现了?</p> <p>Label:体育</p>	<p>Text:</p> <p>In the 36th round of Serie A, Inter reversed Empoli 4-2 at home, Lautaro scored two more goals, including his team's second goal, to equalize the score, and then in the 64th minute, he made the score 3-2, which made a great contribution for Inter to take the game. Now Lautaro is in great form, scoring 11 goals in the last 11 games. In Serie A, he scored 19 goals, which is a new high, and although he ranks third in Serie A, he has fewer penalty kicks, only 3, and both Immobile (27 goals) and Vlahovec (23 goals) have more penalty kicks, scoring 7 in front and 5 behind. So sporting war goals, Lautaro progressed a lot. Lautaro's outbreak for Argentina World Cup is definitely good, there is such a fierce, Messi World Cup trip is stable, right? There is no doubt that Messi has an additional helper. Lautaro, Di Maria, Messi is definitely the top match. There should also be Juve's Dybala, Fiorentina striker Nicolas Gonzalez, Inter Milan's Joaquin Correa, Atletico Madrid's Angel Correa, Sevilla winger Ocampos, and of course, Big Paris' Icardi, which is so enviable to hear. So this World Cup trip, Argentina's front line can absolutely and any team than goals, of course, to take the championship depends on the defense, but can more than the other side to score a goal, it really depends on how the front line led by Lionel Messi performance?</p> <p>Label: Sports</p>

**Figure 3.** Text length distribution of Sohu News dataset.

From the experimental results of Sohu News in Table 4, it can be seen that the classification accuracy of the model proposed in this paper obtained by the convolutional pooling network after fusing the weighted local features with the global features is higher than other models in the case of increasing text length.

Table 4. Experimental results on Sohu News dataset.

Model	Accuracy (%)	Macro-F (%)
TextCNN	92.36	92.32
BiLSTM	92.84	92.91
DPCNN	92.97	92.94
BiLSTM+CNN	94.29	94.33
BERT	94.41	94.42
BT CDC	96.08	96.09

3.3.3. Ablation Experiments Result

To better verify the role of each module in the model, corresponding ablation experiments are designed, where BT CDC-B is the model after ablating global features; BT CDC-ATT is the model after ablating attention weights of local features; BT CDC-MC is the model after ablating multi-level local feature fusion; BT CDC-EP is the model after using general convolution pooling layer instead of equal-length convolution pooling layer. The experimental results are shown in Table 5.

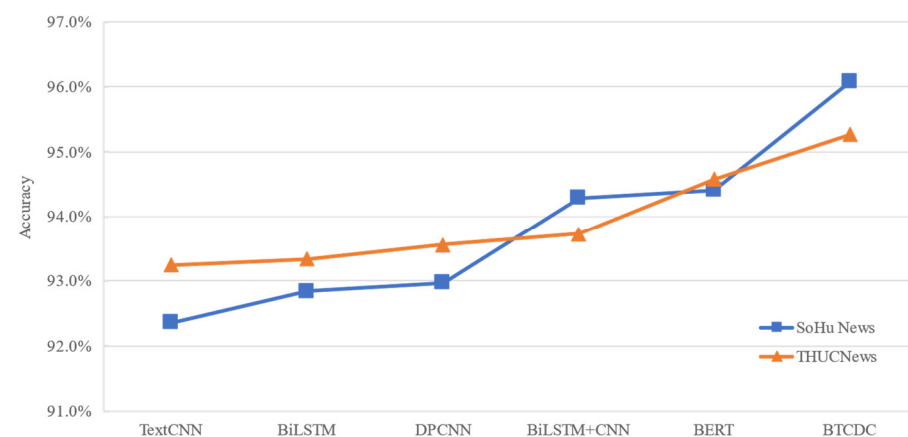
Table 5. Accuracy of each model on the ablation experiment.

Model	THUCNews	SouHuNews
BT CDC	95.27	96.58
BT CDC-ATT	94.12	93.14
BT CDC-MC	94.36	93.51
BT CDC-B	93.85	92.74
BT CDC-EP	94.68	93.89

4. Discussion

4.1. Discussion of Experimental Results on the THUCNews Dataset

From Figure 4, it can be seen that when a single model is used for the classification task, the convolutional pooling of the conventional convolutional neural network and the bi-directional extraction of the BiLSTM perform similarly in the text processing, so it can be seen that the difference between the convolutional neural network and the recurrent neural network classification results is not large, and both local and global features perform well in the text classification task.

**Figure 4.** Accuracy of each model on both datasets.

Compared with the traditional convolutional neural network TextCNN, the deep pyramidal convolutional neural network DPCNN model using deep convolution has a more dominant performance on the classification task; the reason is that its deep convolutional network can better learn the long-range dependencies of text sequences so that the deep

network will perform better in feature extraction. In the paper [42], the deep confidence network in the text classification task significantly outperforms the classical algorithm. And with the BERT training model in the process of powerful pre-training, it can obtain better text semantic information. After fine-tuning the classification task, the combination of a fully connected layer and Softmax classifier can also achieve better results. The reason is that the Transformer structure in the model, combined with the pre-training method and self-attention mechanism, achieves better results in text context semantic learning association. In one of the papers [43], it is again verified that the pre-training approach based on a large corpus can enhance the classification effectiveness by improving the word embedding representation.

By fusing the global feature representation of the BiLSTM with the local features extracted by the convolutional network, the classification results are better compared to other single-feature models. The bidirectional long and short-term memory network can obtain global semantic features and location information of the text sequences. More information in the text can be obtained through the dual feature representation combined with the convolutional network. Research [44] verified the effectiveness of local features with global semantics for the classification task.

The text classification task often has a range effect of both global and local features on the classification effect, and incorporating certain semantic information can improve the classification performance of the text. In this model, after splicing global features with multi-level weighted local features, the global features are obtained while also focusing on some important local features. In addition, the local features contain different multi-level features, which can obtain a richer semantic representation of text features, and combined with deep isometric convolution and pooling, can better capture the text's long-range dependencies. Compared with other comparison models, it achieves better classification results on the THUCNews data set.

4.2. Discussion of Experimental Results on the Sohu News Dataset

From the experimental results of Sohu News in Figure 4, it can be seen that in the experimental comparison of long text selected from the Sohu News public dataset, the experimental results of the single model are somewhat less accurate under the long text category compared to the short text category. The long text sequences cause the model to be more difficult to learn the semantic association of context and not easy to capture more of the long text sequences when convolution pooling to extract features. The BERT model does not differ much from the short text category in terms of its pre-training model and efficient global feature learning.

From the accuracy change trend graph in Figure 5, it can be seen that the accuracy of the model shows a gradual increase after the number of iterations is increased. In the long text classification scenario, the model with the fusion of global features and local features has a better classification effect than in the short text scenario. The model also outperforms other models on the dataset with increasing text sequences, which indicates that the fusion of global feature learning and multi-level feature combination with weighted local features is more effective for the classification of long text sequences. The feature extraction with multi-level convolutional pooling can obtain better feature representation results, which helps to improve the classification performance of the model, and verifies that the fusion of multi-level weighted local features with global features is validated for long text classification tasks.

4.3. Discussion of the Results of Ablation Experiments

From the results of the ablation experiments in Figure 6, it can be seen that the classification effect is significantly lower on vector representations lacking global features. The accuracy rate is significantly lower on long text datasets, indicating that the classification effect is better after fusing global features containing contextual semantics. The attention mechanism can focus more on the important elements in the model and effectively improve

the fused feature representation. In addition, with the addition of multi-level local feature fusion and equal-length convolutional pooling, it can better represent the features and improve classification efficiency.

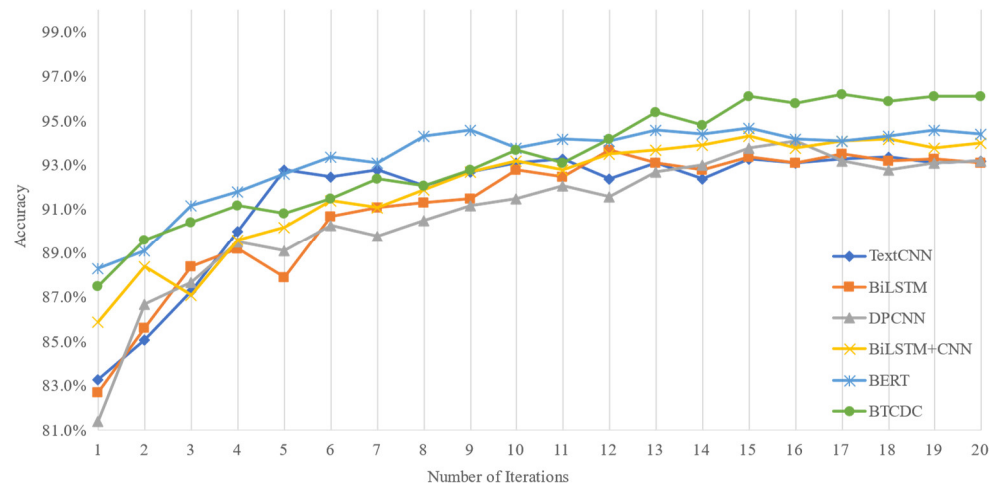


Figure 5. Trend of accuracy on Sohu News dataset.

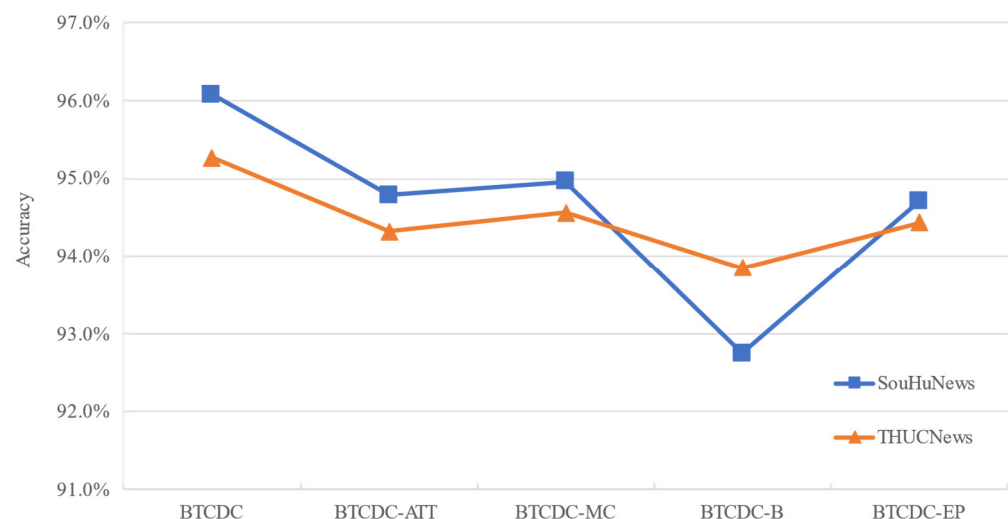


Figure 6. Results of ablation experiments.

5. Conclusions

Currently, the field of text classification has become a popular application in the field of natural language processing, and many scholars have carried out research work on this application area. This paper first summarizes the research progress in the field of text classification in recent years by summarizing some basic models, improvement points, advantages, and disadvantages of the existing research results. Finally, this paper proposes a text classification model based on a pre-trained model with multiple feature fusion, aiming to solve the problems of insufficient feature extraction and information fusion for the field of long text classification. The method firstly inputs word vectors into the BERT model and convolutional neural network in parallel, and the local features combined with different levels of features are obtained by the convolutional neural network, which increases the semantic information representation of features. In addition, the weighted local feature vectors are obtained by combining attention mechanisms, which reduces the influence of multiple features. However, in the local feature extraction of long text, often some information of the original features will also be lost in this process, and further optimization of the attention mechanism in the weight of local features can be tried in the

future. Then the global features obtained by the BERT model are spliced with it to obtain the fused feature representation. The long-range feature representation of text sequences is captured by an equal-length convolutional network, and the contextual semantic global features obtained by the BERT model combined with the weighted multi-level local features can enrich the feature information while effectively avoiding local features from affecting the overall feature performance.

The experimental results show that the proposed model performs well on the publicly available news text dataset, and the classification effect and accuracy are significantly improved compared with other network models. This paper investigates the combination of neural networks and pre-trained models in the field of text classification to further improve the semantic representation of text features. However, due to the difficulty of text information representation and the complexity of text content, information fusion and acquisition are still the focus and difficulty in this field. In the future, we will continue to combine deep learning models in text tasks [45], further investigate enhanced semantic information and better text feature representation methods for long text tasks, the representation of global feature information by pre-trained models with attention mechanisms to represent text semantics, and the combination of external semantic information, etc. to obtain better feature representation.

Author Contributions: Conceptualization, X.Y.; methodology, X.Y.; software, T.Z.; validation, T.Z.; investigation, X.Y.; resources, Y.L.; data curation, L.H.; writing—original draft preparation, T.Z. and L.H.; writing—review and editing, X.Y. and T.Z.; supervision, Y.L.; funding acquisition, L.H. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the research and application of deep learning open sharing platform based on natural language processing (2020YFQ0056).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this paper are open-source which are available at <http://thuctc.thunlp.org/> (accessed on 15 January 2022) and <http://www.sogou.com/labs/> (accessed on 15 January 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Salehan, M.; Kim, D.J. Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decis. Support Syst.* **2016**, *81*, 30–40. [CrossRef]
2. Khatua, A.; Khatua, A.; Cambria, E. A tale of two epidemics: Contextual Word2Vec for classifying twitter streams during outbreaks. *Inf. Processing Manag.* **2019**, *56*, 247–257. [CrossRef]
3. Koleck, T.A.; Dreisbach, C.; Bourne, P.E.; Bakken, S. Natural language processing of symptoms documented in free-text narratives of electronic health records: A systematic review. *J. Am. Med. Inform. Assoc.* **2019**, *26*, 364–379. [CrossRef]
4. Xu, D.; Tian, Z.; Lai, R.; Kong, X.; Tan, Z.; Shi, W. Deep learning based emotion analysis of microblog texts. *Inf. Fusion* **2020**, *64*, 1–11. [CrossRef]
5. Luo, Y.; Xu, X. Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *Int. J. Hosp. Manag.* **2021**, *94*, 102849. [CrossRef] [PubMed]
6. Luhn, H.P. The Automatic Creation of Literature Abstracts. *IBM J. Res. Dev.* **1958**, *2*, 159–165. [CrossRef]
7. McCallum, A.; Nigam, K. A comparison of event models for Naive Bayes text classification. In *AAAI-98 Workshop on Learning for Text Categorization*; AAAI Press: Menlo Park, CA, USA, 1998; pp. 41–48.
8. Zhai, Y.; Song, W.; Liu, X.; Liu, L.; Zhao, X. A chi-square statistics based feature selection method in text classification. In *Proceedings of the 2018 IEEE 9th International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, China, 23–25 November 2018; pp. 160–163.
9. Liu, C.-z.; Sheng, Y.-x.; Wei, Z.-q.; Yang, Y.-Q. Research of text classification based on improved TF-IDF algorithm. In *Proceedings of the 2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, Lanzhou, China, 24–27 August 2018; pp. 218–222.
10. Han, K.-X.; Chien, W.; Chiu, C.-C.; Cheng, Y.-T. Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter DataSet. *Appl. Sci.* **2020**, *10*, 1125. [CrossRef]

11. Hinton, G.E.; Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. *Science* **2006**, *313*, 504–507. [CrossRef]
12. Sutskever, I.; Martens, J.; Hinton, G.E. Generating text with recurrent neural networks. In Proceedings of the 2011 International Conference on Machine Learning (ICML), Bellevue, WA, USA, 28 June–1 July 2011.
13. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
14. Elnagar, A.; Al-Debsi, R.; Einea, O. Arabic text classification using deep learning models. *Inf. Processing Manag.* **2020**, *57*, 102121. [CrossRef]
15. Kim, Y. Convolutional Neural Networks for Sentence Classification. *arXiv* **2014**, arXiv:1408.5882.
16. Johnson, R.; Zhang, T. Deep pyramid convolutional neural networks for text categorization. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 30 July–4 August 2017; Volume 1: Long Papers, pp. 562–570.
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
18. Pota, M.; Esposito, M.; De Pietro, G.; Fujita, H. Best Practices of Convolutional Neural Networks for Question Classification. *Appl. Sci.* **2020**, *10*, 4710. [CrossRef]
19. Shu, Z.; Zheng, D.; Hu, X.; Ming, Y. Bidirectional Long Short-Term Memory Networks for Relation Classification. In Proceedings of the 29th Pacific Asia Conference on Language, Information and Computation, Shanghai, China, 30 October–1 November 2015.
20. Xu, C.; Huang, W.; Wang, H.; Wang, G.; Liu, T.-Y. Modeling local dependence in natural language with multi-channel recurrent neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 5525–5532.
21. Du, J.; Vong, C.-M.; Chen, C.P. Novel efficient RNN and LSTM-like architectures: Recurrent and gated broad learning systems and their applications for text classification. *IEEE Trans. Cybern.* **2020**, *51*, 1586–1597. [CrossRef]
22. Zhou, P.; Shi, W.; Tian, J.; Qi, Z.; Li, B.; Hao, H.; Xu, B. Attention-based bidirectional long short-term memory networks for relation classification. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; Volume 2: Short Papers, pp. 207–212.
23. Zheng, W.; Liu, X.; Yin, L. Sentence Representation Method Based on Multi-Layer Semantic Network. *Appl. Sci.* **2021**, *11*, 1316. [CrossRef]
24. Zheng, W.; Yin, L. Characterization inference based on joint-optimization of multi-layer semantics and deep fusion matching network. *PeerJ Comput. Sci.* **2022**, *8*, e908. [CrossRef] [PubMed]
25. Zheng, W.; Zhou, Y.; Liu, S.; Tian, J.; Yang, B.; Yin, L. A deep fusion matching network semantic reasoning model. *Appl. Sci.* **2022**, *12*, 3416. [CrossRef]
26. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems 30: Annual Conference, Long Beach, CA, USA, 4–9 December 2017.
28. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.
29. Lan, Z.; Chen, M.; Goodman, S.; Gimpel, K.; Sharma, P.; Soricut, R. Albert: A lite bert for self-supervised learning of language representations. *arXiv* **2019**, arXiv:1909.11942.
30. Martinčić-Ipšić, S.; Miličić, T.; Todorovski, L. The Influence of Feature Representation of Text on the Performance of Document Classification. *Appl. Sci.* **2019**, *9*, 743. [CrossRef]
31. Yang, L.; Hongbin, D. Text sentiment analysis based on feature fusion of convolution neural network and bidirectional long short-term memory network. *J. Comput. Appl.* **2018**, *38*, 3075.
32. Zhang, J.; Liu, F.A.; Xu, W.; Yu, H. Feature Fusion Text Classification Model Combining CNN and BiGRU with Multi-Attention Mechanism. *Future Internet* **2019**, *11*, 237. [CrossRef]
33. Luo, X.; Wang, X. Research on multi-feature fusion text classification model based on self-attention mechanism. *J. Physics: Conf. Ser.* **2020**, *1693*, 012071. [CrossRef]
34. Xie, J.; Hou, Y.; Wang, Y.; Wang, Q.; Li, B.; Lv, S.; Vorotnitsky, Y.I. Chinese text classification based on attention mechanism and feature-enhanced fusion neural network. *Computing* **2020**, *102*, 683–700. [CrossRef]
35. Jang, B.; Kim, M.; Harerimana, G.; Kang, S.; Kim, J.W. Bi-LSTM Model to Increase Accuracy in Text Classification: Combining Word2vec CNN and Attention Mechanism. *Appl. Sci.* **2020**, *10*, 5841. [CrossRef]
36. Abdi, A.; Shamsuddin, S.M.; Hasan, S.; Piran, J. Deep learning-based sentiment classification of evaluative text based on Multi-feature fusion. *Inf. Processing Manag.* **2019**, *56*, 1245–1259. [CrossRef]
37. Sun, M.; Li, J.; Guo, Z.; Yu, Z.; Zheng, Y.; Si, X.; Liu, Z. Thuctc: An Efficient Chinese Text Classifier. GitHub Repos. 2016. Available online: <https://github.com/thunlp/THUCTC> (accessed on 15 January 2022).
38. Wang, C.; Zhang, M.; Ma, S.; Ru, L. Automatic online news issue construction in web environment. In Proceedings of the 17th International Conference on World Wide Web, Beijing, China, 21–25 April 2008; pp. 457–466.

39. Mao, X.; Chang, S.; Shi, J.; Li, F.; Shi, R. Sentiment-Aware Word Embedding for Emotion Classification. *Appl. Sci.* **2019**, *9*, 1334. [\[CrossRef\]](#)
40. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
41. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
42. Jiang, M.; Liang, Y.; Feng, X.; Fan, X.; Pei, Z.; Xue, Y.; Guan, R. Text classification based on deep belief network and softmax regression. *Neural Comput. Appl.* **2018**, *29*, 61–70. [\[CrossRef\]](#)
43. Rezaeinia, S.M.; Rahmani, R.; Ghodsi, A.; Veisi, H. Sentiment analysis based on improved pre-trained word embeddings. *Expert Syst. Appl.* **2019**, *117*, 139–147. [\[CrossRef\]](#)
44. Liu, G.; Guo, J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing* **2019**, *337*, 325–338. [\[CrossRef\]](#)
45. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning-based text classification: A comprehensive review. *ACM Comput. Surv.* **2021**, *54*, 1–40. [\[CrossRef\]](#)