Preprints.org

Article

Research on the Classification of Work Orders based on BERT and Feature Fusion

Xiong Yun Peng , Lian Guo Chen [*] , Kuo Jun Cao [*]

Article

Research on the Classification of Work Orders based on BERT and Feature Fusion

Yun Peng Xiong [1], Guolian Chen [2],* and Junkuo Cao [3],*

[1]   School of Information Science and Technology, Hainan Normal University, HaiKou, 571127, China;
      a18579118404@163.com

[2]   State-owned Assets Management Office, Hainan Normal University, HaiKou, 571158, China;
      cgl@hainnu.edu.cn

[3]   Information Network and Data Center, Hainan Normal University, HaiKou, 571158, China;
      cjk@hainnu.edu.cn

*   Correspondence: cgl@hainnu.edu.cn; cjk@hainnu.edu.cn

Abstract: Convolutional neural networks (CNNs) face challenges in capturing long-distance text correlations, and Bidirectional Long Short-Term Memory (BiLSTM) networks exhibit limited feature extraction capabilities for text classification of work order. To address the abovementioned problems, this work utilizes an ensemble learning approach to integrate model elements efficiently. This study presents a method for classifying work order texts using a hybrid neural network model called BERT-BiLSTM-CNN. First, use Bert for preprocessing to obtain text vector representations. Then, capture context and process sequence information through BiLSTM. Next, capture local features in the text through CNN.   Finally, obtain classification results through Softmax. Through comparative analysis, the method of fusing these three models is superior to other hybrid neural network model architectures in multiple classification tasks. It has a significant effect on work order classification.

1. Introduction

Within natural language processing, text classification is a widely used field. Some conventional machine learning techniques, including Naive Bayes [1], Decision Tree [02], and K Nearest Neighbor (KNN) [3,4], etc. Although these methods can achieve specific classification purposes, they also have limitations such as weak computing power, low efficiency, and poor performance. With the deepening of deep learning, compared to conventional machine learning techniques, its classification approaches yield a far higher classification effect. The commonly used model is the Convolutional Neural Network (CNN) used for text [5,6], Recurrent Neural Network (RNN) [7,8], Long Short-Term Memory network (LSTM) [9], Bi-directional Long Short-Term Memory (BiLSTM)[10], and Attention Mechanism [11]. However, these models still need to improve at capturing complex semantic and syntactic information in text and must be based on the word vector Word2Vec [12]. Because Word2Vec still has a lot of semantic ambiguity, it cannot capture accurate semantic information for some words with multiple meanings [13]. The pre-trained model BERT (Bidirectional Encoder Representations from Transformers, BERT) [14], developed by the Google team in 2018, pertinently solves the semantic ambiguity problem. Studies have shown that in natural language processing tasks like text categorization and sentiment analysis, Bert outperforms Word2Vec [13,15,25,27].

Despite significant advancements in deep and machine learning, a single model frequently has drawbacks, such as a lengthy training period, limited scope, large number of parameters, and poor interpretability in completing text classification tasks, and its obtained results are often difficult to meet the ideal expectations. On the one hand, the information provided by the word vectors obtained from a single model is limited, and it's also important to take word relationships into account, as well

as the position information and context information of words in sentences or documents. On the other hand, researchers generally find that a hybrid neural network composed of multiple neural network models organically fused [16,17] can usually explore the advantages of a single model while overcoming its disadvantages and improving the classification performance to a large extent. Given the abovementioned circumstances, this study proposes a BERT-BiLSTM-CNN (BBLC) hybrid neural network model, which can effectively improve work order text categorization performance by integrating current models.

2. Related Work

Researchers have been more concerned about the hybrid neural network model in recent years. For example, Li [18] et al. used BERT combined with a full connection layer to conduct sentiment analysis on stock comments. Compared with CNN, RNN, BiLSTM, and other models, BERT combined with a full connection layer has a better effect in sentiment classification. Ren [19] and Li [20] et al. combined BERT and BiLSTM models for emotion analysis and the fusion model performs well in the classification of irregular texts' emotion analysis. Kaur [21] et al., by combining a CNN with the BERT model, discovered that the CNN model could extract a more significant amount of semantic information, improving the accuracy of the classification model. Xie [22] and Deng [23] et al. presented a feature enhancement fusion model that utilizes an attention mechanism to integrate deep learning models—like CNN and LSTM—and enhances the capture of keyword data. Zhao [24] et al. use BiLSTM-CNN, which integrates multiple features, to perform community Q&A text matching. After multi-feature fusion, the model has a more robust matching performance. Bao [25] et al. proposed a hybrid BERT-based short-text classification algorithm that blends CNN and BiGRU, which is significantly superior to other single models in short-text classification tasks. Jiang [26] and Kaur [27] et al. combined BiLSTM and CNN based on BERT and achieved good classification results in both binary classification and multi-classification tasks. Based on the above analysis, this paper aims to optimize the utilization of BERT, BiLSTM, and CNN by integrating these models and proposing a text classification approach for work orders using a mixed neural network model called BBLC.

3. Model Design

3.1. BERT

The two stages of the BERT paradigm are called fine-tuning and pre-training. A multi-layer bidirectional Transformer encoder layered with 12 encoder structures makes up its fundamental construction. Each Encoder structure has 12 self-attention heads and 768 hidden units per feedforward layer. The output representation of each word is obtained by summing three Embedding vectors: Token Embedding, Position Embedding, and Segment Embedding. Figure 1 displays the BERT input feature diagram. Two tasks are used to pre-train the model: Next Sentence Prediction (NSP) and Masked Language Modelling (MLM). The words in the input text are randomly sheltered by MLM, which enables the model to anticipate the word's original form, thereby learning the word in the text to be expressed in the context.NSP is used to train models to understand the order and relationship between sentences, predicting whether two sentences are consecutive. The fine adjustment is simple, mainly by freezing most parameters, only for parameters requiring fine-tuning.
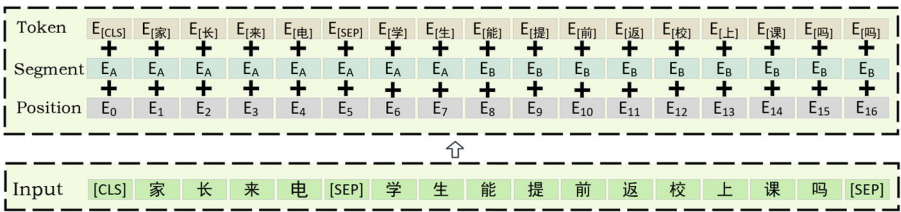

Figure 1. BERT feature input diagram.

3.2. BiLSTM

BiLSTM is a combination of forward-to-the-clock LSTM and back-to-the-clock, one of which is a forward processing sequence, and the other is a reverse processing sequence. By stacking two layers of LSTM, the model is freed from the constraint of predicting the next moment's output solely based on the temporal information from the previous moment. In the BBLC model, the BiLSTM layer processes the output of the BERT, extracting and encoding the above-text information and semantic characteristics of the text. Figure 2 displays the structural diagram of BiLSTM.
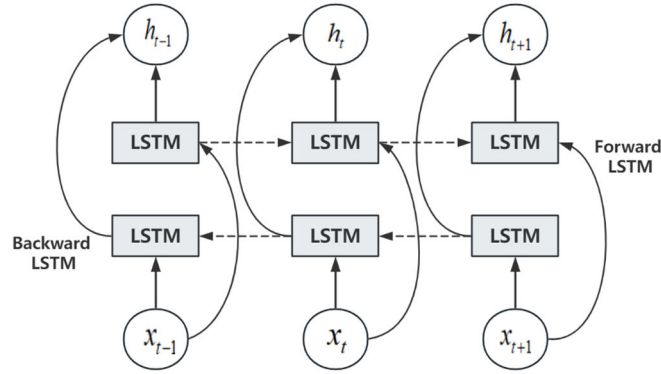

Figure 2. BiLSTM model structure diagram.

After processing is completed, the outputs of the two LSTMs are stacked together as the current output of the BiLSTM. The calculation formula is (1) to (3).

$$\overrightarrow{h_t} = L\overrightarrow{STM}\ (x_t, \overrightarrow{h}_{t-1}\ ) \tag{1}$$

$$\overleftarrow{h_t} = L\overleftarrow{STM}\ (x_t, \overleftarrow{h}_{t-1}\ ) \tag{2}$$

$$h = [\overrightarrow{h_t} \oplus \overleftarrow{h_t}] \tag{3}$$

In formulas (1) to (3), the variable x_t denotes the input vector for time t. $\overrightarrow{h}_{t-1}$ denotes the forward hidden layer vector at time $t-1$, while $\overleftarrow{h}_{t-1}$ denotes the reverse hidden layer vector at time $t-1$. Formula (3) to concatenate the two outputs, $\oplus$ represents the concatenation symbol.

3.3. CNN

A pooling layer and a convolution layer make up the CNN layer. The convolution kernel is applied to all possible subsequences in the text, allowing for the effective capture of local patterns and features in the input text. This process generates feature maps that correlate to these patterns and features. The pooling layer decreases the dimensionality of the feature map while retaining the most salient characteristics. One often used pooling method is max pooling, which outputs the highest value from each feature vector.

$$Y[i,j] = \max_{m,n} X[i \times k + m, j \times k + n] \tag{4}$$

Slide a window of size $k \times k$ with step s in the input feature map, and assign the highest value in each window as the value of the associated point in the output feature map. In formula (4), $Y[i,j]$ represents the element in row i and column j of the output feature map, $X[i \times k + m, j \times k + n]$ represents the

element that enters the polarized window position in the characteristic diagram, m and n are the relative positions in each window, with values ranging from 0 to $k-1$.

3.4. BERT-BiLSTM-CNN

BiLSTM-CNN is the downstream module, and BERT is the upstream module in the BBLC model architecture. Figure 3 shows the approximate working process of the BBC model:

(1) First, process the data set. $W_1, W_2...W_n$ is the input text in Figure 3. The input text is converted into word vectors $T_1, T_2...T_n$ after the Transformer encoder, which can be trained as model parameters.

(2) BERT output continues to feed into a two-way LSTM network. BiLSTM uses forward and backward LSTM units for bidirectional modeling of input sequences. The information included in the input sequence is captured by the hidden states of each time step, which are the BiLSTM outputs. CNN receives the spliced forward and backward outputs of the BiLSTM.

(3) Multiple convolution kernels of the CNN convolution layer extract local features by sliding over the input data. The pooling layer receives local feature maps from the convolutional layer as input and performs max pooling operations for each feature map.

(4) The fully connected layer receives the feature representation from the BiLSTM-CNN module and obtains the output y through the linear transformation of the weight matrix. Finally, the Softmax function receives the output of the full connection layer and processes it indexically and normalized to obtain a probability p_i for each category. The calculation formula is (5)-(6).

$$y = f(W * x + b) \tag{5}$$

$$p_i = \frac{e^{y_i}}{\sum_{j=1}^{n} e^{y_j}} \tag{6}$$

Formula (5): f represents the activation function; W represents the weight matrix, x is the input eigenvector, and b is the bias value; Formula (6): y_i represents the score of the i category, and n represents a total of n categories.
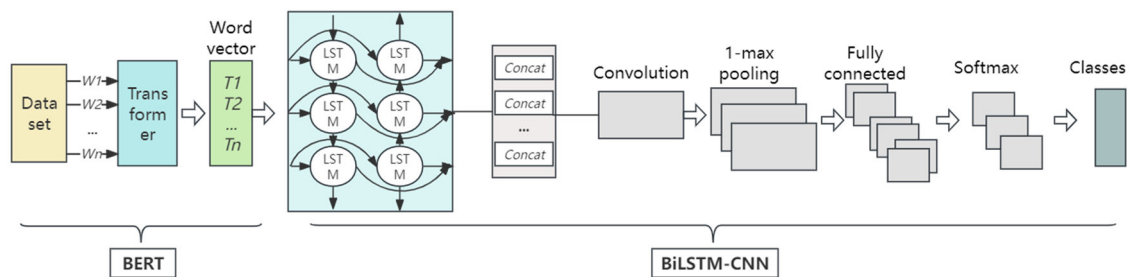

Figure 3. BERT-BiLSTM-CNN model overall frame diagram.

4. Experiments and Results

4.1. Data Set

The experimental data is based on the Open Data Innovation Application Competition of Jiangxi Province and comes from the work order data of Xinyu City citizens' petitioning in 2021. There are 72,300 pieces of the work order data set and 52,200 pieces of data after annotation and cleaning. The training set and the test set were separated from the data set using a ratio of 0.8:0.2. There are four categories: consultation, Complaint, Help, and Advice. Table 1 displays the data distribution.

Table 1. Dataset distribution.

Label	Text category	Training set	Test set
0	Consult(咨询)	14260	3843
1	Complain(投诉)	16246	4319
2	Seek Assist(求助)	9669	2549
3	Suggestion(建议)	1207	325

To facilitate training, the labels for Consultation, Complaint, Help, and Advice are 0, 1, 2, and 3, respectively. Table 2 shows part of the data set.

Table 2. Partial data of dataset

Label	Text content
0	市民来电咨询：广东省河源市回余人员有什么疫情防控措施？ Citizens call for advice: What epidemic prevention and control measures do people from Heyuan City, Guangdong Province return to Xinyu City？
1	新余四中不经过家长同意补课收取补课费 不允许家长反抗 说一句不同意都不行 把孩子教好是教师应尽的责任与义务 要有高效课堂 减轻孩子负担 而不是靠收费和补课来加重孩子和家长的压力。 Xinyu No. 4 Middle School does not charge remedial fees without parents' consent. Parents are not allowed to resist and say that they do not agree. It is the responsibility and obligation of teachers to teach children well. It is necessary to have efficient classrooms to reduce the burden on children rather than rely on fees and remedial classes to increase the pressure on children and parents.
2	市民来电：高速爆胎的求助。 Public call: high-speed flat tire for help
0	近日省教育厅发文称全省2021届初三学生中考改革，但我去省教育厅却说以当地中招政策为主。请问这届初三学生中考是否改革？ Recently, the provincial Department of Education issued a document saying that the province's 2021 junior high school examination reform, but I went to the provincial Department of Education said that the local recruitment policy is based. May I ask whether the junior high school entrance examination is reformed?
3	吴先生来电建议减少南源路货车通行量。 Mr. Wu called to suggest reducing truck traffic on Nanyuan Road.
1	市民来电反映厦门某宝马4S店退回定金且不售卖宝马车，认为不合理! The public call reflects that a BMW 4S store in Xiamen returns the deposit and does not sell BMW, which is unreasonable!
2	市民来电：虎山路长青小学（市民表示不愿意透露年级）下完延时课下午5点以后是否可以由学校老师（带去其他教学地方进行教学（教学内容是课堂学习内容），市民表示是普遍现象。 Call from the public: Whether the school teacher can take the delayed class to other teaching places (the teaching content is the classroom learning content) after 5pm after Hushan Road Changqing Primary School (the public said that it is a common phenomenon.

4.2. Experimental Environment and Parameters Design

All experiments were performed in GeForce RTX 4060 and Inter(R) Core(TM) i7-12650 configurations. The experimental framework is built with PyTorch (2.1.2+cu118) and runs with Python3.8 in the Anaconda environment. The best-combined effect was obtained by training 15 epochs. With Dropout set to 0.5, turning off 50 percent of neurons during training not only effectively reduces overfitting but also reduces coupling between neurons, and the model achieves convergence with greater ease. The model's Learning_rate is 0.001, and the Batch_size is 128. The larger the Batch size is, the training speed of the model will be slowed down, and the gradient oscillation will be severe, which is not conducive to convergence. In the BiLSTM network, Hidden size refers to the dimension size of the hidden state in each LSTM cell. After multiple rounds of parameter tuning, the

Hidden_size of 768 for the model was found to be relatively simple and can effectively reduce overfitting. Due to the limited length of the work order text, the maximum height (Max_length) is set to 300. Setting three different sizes of convolutional kernels can help the model better capture semantic information at various scales, thus enhancing its expressive power and generalization performance. Table 3 shows the specific settings of the experimental parameters.

Table 3. Experimental parameters.

Hyperparameter	Value	Hyperparameter	Value
Epoch	15	Optimizer	Adam
Dropout	0.5	Hidden_size	768
Learning rate	10-3	Max_length	300
Batch_size	128	Kernel_sizes	[7–9]

4.3. Experimental Comparison and Result Analysis

4.3.1. Model Comparison

To verify the robustness of the BBLC model, the accuracy, precision, recall, and F1 score on the test set were compared across various models using multiple sets of experiments. Table 4 compares the experimental results of the BBLC model with other models on the test set.

Table 4. Comparison of experimental results.

Model	Accuracy	Precision	Recall	F1-score
BERT-CNN	0.950 9	0.935 4	0.950 3	0.939 5
BERT-BiGRU	0.934 5	0.940 5	0.912 6	0.921 3
BERT-BiLSTM	0.930 6	0.908 0	0.927 0	0.912 9
BERT-BiGRU-CNN	0.942 0	0.938 5	0.934 4	0.932 2
BERT-BiLSTM-CNN	**0.953 6**	**0.954 1**	**0.951 9**	**0.950 9**

* The bold text is the model of the article.

The experiment uses the F1 score as the primary evaluation metric, which simultaneously considers the model's precision and recall. According to Table 4, it can be observed that, compared to other models, the BBC model has the highest F1 score. A high F1 score indicates that the model has excellent performance in classification tasks and can effectively distinguish and predict various categories. Second, the recall rate is also the highest among them, and a high recall rate usually indicates that the model can find more correct predictions in real positive samples. Thirdly, even when the recall is at its highest, the model also exhibits the highest precision. This indicates that the model, when predicting positive samples, can minimize the misclassification of negative samples as positive, demonstrating better precision. Fourth, the high accuracy of a model indicates that the model can correctly predict more results in all samples and has better overall forecasting ability. These four highest metrics imply that the BBLC model has the most optimal overall performance. Figure 4 shows the accuracy and loss curves of the BBLC model on the training and validation sets.
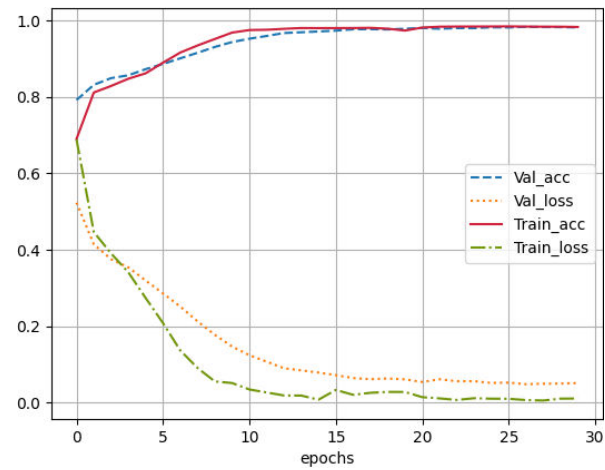
Figure 4. Training set accuracy and loss curves.

The confusion matrix in Figure 5 illustrates the classification performance of different models on the test set. According to the confusion matrix, the BERT-CNN model correctly classified 10,466 data points, Bert-BiLSTM classified 10,242, BERT-BiGRU-CNN classified 10,368, while the article model BBLC classified 10,497, which exhibits the best performance among these four main models. Although the BBLC model misclassifies most 1-class labels into 2-class labels, its overall classification ability is the best, achieving the highest accuracy in ticket text classification.
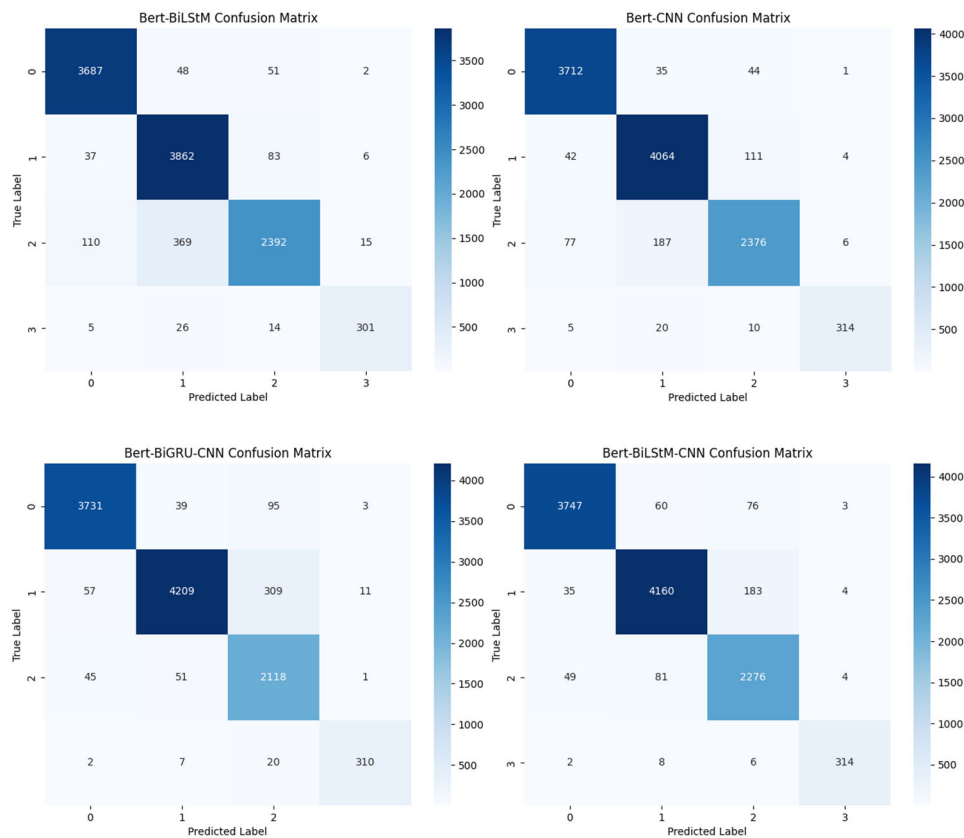

Figure 5. Confusion matrices for different models.

The attention mechanism in a hybrid neural network model usually allows the model to make greater use of features that are retrieved from different components. There are certain variations, though, because work order text is distinct. To evaluate the effectiveness of the Attention mechanism in the work order text classification task, the Attention mechanism is integrated into the hybrid neural

network for comparison. As shown in Table 5, when the Attention mechanism is incorporated into the models, F1 scores of the models are improved; Table 6 shows that when incorporating the Attention mechanism into hybrid models like CNN, the precision is somewhat enhanced by the influence of recall. However, the overall metric F1 score decreases; this is because when Attention is combined with CNN, the Attention mechanism will cause the model to overlook other significant aspects in favor of paying excessive attention to local features, resulting in some helpful information being ignored, thus affecting the performance of the model.

Table 5. Attention but no CNN models comparison.

Model	Accuracy	Precision	Recall	F1-score
BERT-BiGRU	0.934 5	**0.940 5**	0.913 6	0.922 3
BERT-BiGRU-Attention	**0.946 9**	0.938 6	**0.944 8**	**0.938 7**
BERT-BiLSTM	0.930 6	0.908 0	0.927 0	0.912 9
BERT-BiLSTM-Attention	**0.939 1**	**0.938 2**	**0.930 9**	**0.931 4**

Table 6. Attention and CNN models comparison

Model	Accuracy	Precision	Recall	F1-score
BERT-BiGRU-CNN	**0.942 0**	**0.938 5**	**0.934 4**	**0.932 2**
BERT-BiGRU-Attention-CNN	0.929 9	0.927 4	0.906 2	0.912 5
BERT-CNN	**0.950 9**	0.935 4	**0.950 3**	**0.939 5**
BERT-Attention-CNN	0.944 6	**0.944 0**	0.928 7	0.933 5
BERT-BiLSTM-CNN	**0.953 6**	**0.954 1**	**0.951 9**	**0.950 9**
BERT-BiLSTM-CNN-Attention	0.950 5	0.952 8	0.945 7	0.947 0

4.3.2. Comparison of the Classification Effects of Different Parameters

In the fusion model BBLC, the hyperparameter setting is essential. The different hyperparameter settings of each model result in different classification effects. For example, different Hidden sizes in BiLSTM lead to different complexity of model learning sequence features, thus affecting the ability to model sequence data. The various sizes of Convolutional kernels in CNN will impact the model's capacity to extract features from text segments of diverse lengths. To further enhance the performance of the BBLC, experiments involving the size of Convolutional kernels and the dimensions of hidden sizes were conducted.

Table 7 presents the scenario where the convolutional kernel size varies while other parameters remain constant. Table 8 shows that as the convolutional kernel size increases, the model's recall and F1 score also improve. When the convolutional kernel is too large, the model tends to overfit. Overfitting occurs when the model becomes overly complex and excessively adapts to the characteristics of the training data, leading to a loss of generalization ability. So, choose the appropriate convolution kernel size; Table 8 shows the dimensions of different hidden layers when other parameters are the same. Too small a hidden layer will lead to poor performance of the model on the test set, inadequate dimensionality of the hidden layer will result in subpar performance of the model on the test dataset, and an excessively large hidden layer can lead to overfitting. Therefore, after a series of comparative experiments, I selected Num_layer=1, kernel_sizes=[7–9], and Hidden_sizes=768 for hyperparameter training.

Table 7. Experimental comparison of Convolution Kernel sizes.

Model			BERT-BiLSTM-CNN			
Num_layer	Kernel_sizes	Hidden_sizes	Accuracy	Precision	Recall	F1-score
1	[2,3,4]	768	0.946 1	0.945 0	0.945 4	0.943 1
1	[3,4,5]	768	0.949 1	0.951 0	0.941 4	0.944 2
1	[4,5,6]	768	0.946 7	0.936 0	0.946 1	0.937 8
1	[5,6,7]	768	0.937 7	0.937 7	0.948 5	0.939 2

1	[6,7,8]	768	0.952 9	0.952 1	0.949 5	0.949 0
1	[7,8,9]	768	**0.953 6**	**0.954 1**	0.951 9	**0.950 9**
1	[8,9,10]	768	0.946 4	0.945 9	0.947 4	0.943 7
1	[9,10,11]	768	0.952 7	0.953 2	0.949 5	0.949 1
1	[10,11,12]	768	0.951 9	0.950 9	**0.956 3**	0.948 7

Table 8. Experimental comparison of Hidden sizes.

Model			BERT-BiLSTM-CNN			
Num_layer	Kernel_sizes	Hidden_sizes	Accuracy	Precision	Recall	F1-score
1	[7,8,9]	128	0.937 9	0.944 0	0.933 7	0.935 3
1	[7,8,9]	256	0.948 3	0.953 0	0.943 4	0.945 9
1	[7,8,9]	512	0.952 1	0.952 8	0.946 9	0.948 0
1	[7,8,9]	768	**0.953 6**	**0.954 1**	**0.951 9**	**0.950 9**

5.Conclusions

To enhance the efficiency of retrieving the wording of the work order requested by residents, this paper classifies the data of the work order petitioned by citizens. It makes an in-depth analysis of the text content of the work order. In this text classification task, a hybrid neural network classification model (BERT-BiLSTM-CNN) is established, significantly improving the classification performance and thus improving the retrieval efficiency and speed. Firstly, BERT is used to encode the text to obtain rich semantic information and generate word vectors. BiLSTM then models bidirectional sequences of the encoded text to capture sequence information in the text. Then, CNN extracts local features to enhance the model's perception of essential features in the text. After being processed by BiLSTM and CNN, The fully connected layers merge their output features and input the merged characteristics into the Softmax classifier for classification. The hybrid neural network BBLC combines BERT, BiLSTM, and CNN by comprehensively leveraging their advantages in semantic understanding, sequence modeling, and feature extraction, enhancing the efficiency in jobs involving text classification. In our subsequent research, we will further investigate the BERT-BiLSTM-CNN model and the interplay between its components to strengthen its classification performance in future text classification or sentiment analysis tasks.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

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

References

1. Guan G, Guo J, Wang H. Varying Naïve Bayes models with applications to classification of chinese text documents[J]. Journal of Business & Economic Statistics, 2014, 32(3): 445-456.
2. Moraes R, Valiati J F, Neto W P G O. Document-level sentiment classification: An empirical comparison between SVM and ANN[J]. Expert Systems with Applications, 2013, 40(2): 621-633.
3. Jiang S, Pang G, Wu M, et al. An improved K-nearest-neighbor algorithm for text categorization[J]. Expert Systems with Applications, 2012, 39(1): 1503-1509.
4. Bilal M, Israr H, Shahid M, et al. Sentiment classification of Roman-Urdu opinions using Naïve Bayesian, Decision Tree and KNN classification techniques[J]. Journal of King Saud University-Computer and Information Sciences, 2016, 28(3): 330-344.

5.   Chen Y. Convolutional neural network for sentence classification[D]. University of Waterloo, 2015.

6.   Soni S, Chouhan S S, Rathore S S. TextConvoNet: A convolutional neural network based architecture for text classification[J]. Applied Intelligence, 2023, 53(11): 14249-14268.

7.   Lai S, Xu L, Liu K, et al. Recurrent convolutional neural networks for text classification[C]//Proceedings of the AAAI conference on artificial intelligence. Chinese Academy of Sciences:NLPR.2015, 29(1).

8.   Yin W, Kann K, Yu M, et al. Comparative study of CNN and RNN for natural language processing[J]. arXiv preprint arXiv:1702.01923, 2017.

9.   AR D,Manju S B,Songbo T , et al.LSTM Based Text Classification[J].IITM Journal of Management and IT,2021,12(1):62-65.

10.   Liu G, Guo J. Bidirectional LSTM with attention mechanism and convolutional layer for text classification[J]. Neurocomputing, 2019, 337: 325-338.

11.   Galassi A, Lippi M, Torroni P. Attention in natural language processing[J]. IEEE transactions on neural networks and learning systems, 2020, 32(10): 4291-4308.

12.   Zhang D, Xu H, Su Z, et al. Chinese comments sentiment classification based on word2vec and SVMperf[J]. Expert Systems with Applications, 2015, 42(4): 1857-1863.

13.   Shen Y, Liu J. Comparison of text sentiment analysis based on bert and word2vec[C]//2021 IEEE 3rd international conference on frontiers technology of information and computer (ICFTIC). IEEE, 2021: 144-147.

14.   Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

15.   Kale A S, Pandya V, Di Troia F, et al. Malware classification with word2vec, hmm2vec, bert, and elmo[J]. Journal of Computer Virology and Hacking Techniques, 2023, 19(1): 1-16.

16.   Li X, Cui M, Li J, et al. A hybrid medical text classification framework: Integrating attentive rule construction and neural network[J]. Neurocomputing, 2021, 443: 345-355.

17.   Hernández G, Zamora E, Sossa H, et al. Hybrid neural networks for big data classification[J]. Neurocomputing, 2020, 390: 327-340.

18.   Li M, Chen L, Zhao J, et al. Sentiment analysis of Chinese stock reviews based on BERT model[J]. Applied Intelligence, 2021, 51: 5016-5024.

19.   Ren C ,Bin Q ,Yangken C , et al.Sentiment Analysis About Investors and Consumers in Energy Market Based on BERT-BiLSTM[J].IEEE ACCESS,2020,8171408-171415.

20.   Li X, Lei Y, Ji S. BERT-and BiLSTM-based sentiment analysis of online Chinese buzzwords[J]. Future Internet, 2022, 14(11): 332.

21.   Kaur K, Kaur P. BERT-CNN: improving BERT for requirements classification using CNN[J]. Procedia Computer Science, 2023, 218: 2604-2611.

22.   Xie J, Hou Y, Wang Y, et al. Chinese text classification based on attention mechanism and feature-enhanced fusion neural network[J]. Computing, 2020, 102: 683-700.

23.   Deng J, Cheng L, Wang Z. Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification[J]. Computer Speech & Language, 2021, 68: 101182.

24.   Li Z, Yang X, Zhou L, et al. Text matching in insurance question-answering community based on an integrated BiLSTM-TextCNN model fusing multi-feature[J]. Entropy, 2023, 25(4): 639.

25.   Bao T, Ren N, Luo R, et al. A BERT-based hybrid short text classification model incorporating CNN and attention-based BiGRU[J]. Journal of Organizational and End User Computing (JOEUC), 2021, 33(6): 1-21.

26.   Jiang X, Song C, Xu Y, et al. Research on sentiment classification for netizens based on the BERT-BiLSTM-TextCNN model[J]. PeerJ Computer Science, 2022, 8: e1005.

27.   Kaur K, Kaur P. Improving BERT model for requirements classification by bidirectional LSTM-CNN deep model[J]. Computers and Electrical Engineering, 2023, 108: 108699.