Online-Journals.org

iJET International Journal of Emerging Technologies in Learning

iJET | eISSN: 1863-0383 | Vol. 18 No. 21 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijet.v18i21.37843

PAPER

A CNN-Bi-LSTM Model for MOOC Forum Post Classification

Qiaorong Zhang[1](✉), **Lin Sun**[2]

[1]College of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou, China

[2]School of Economics and Management, China University of Petroleum, Beijing, China

zqrzqh@126.com

ABSTRACT

The discussion forum of the massive open online course (MOOC) is a platform for students to communicate with teachers, teaching assistants, and platform managers. It is one of the important factors related to course quality. A reasonable classification of discussion posts in the forum will help students better communicate and solve problems, so as to improve the quality of teaching. Aiming at the classification of discussion forum posts, this paper proposes a text classification model integrating convolutional neural networks (CNN) and bidirectional long-short-term memory (Bi-LSTM). Firstly, the user types and behavior characteristics are analyzed to build the taxonomy. The taxonomy includes three categories: course related, teacher related and platform related. Then, a text classification model is constructed based on CNN and Bi-LSTM. In order to verify the effectiveness of the proposed model, it is applied to the classification of 19285 discussion posts from the MOOC platform of icourse163.org. The overall classification accuracy of the proposed model is 93.6%, which is 12%, 10%, and 8% higher than traditional machine learning methods, CNN and Bi-LSTM, respectively. The model is used for automatic text classification in MOOC discussion forum, which can provide effective help and support for learners, teachers and platform managers, and improve the automation level of MOOC platform.

KEYWORDS

text classification, massive open online courses (MOOC), discussion forum, classification model integrating (CNN), bidirectional long-short term memory (Bi-LSTM)

1 INTRODUCTION

Massive open online courses (MOOCs) are highly praised and welcomed by the majority of learners because of the wide variety of courses, rich content, free form, and openness. In order to meet the personalized learning needs of learners and improve the service level of the platform, most MOOC platforms provide discussion forums to enhance learners' learning experiences in peer interaction, resource retrieval, problem discussion, and so on [1]. The MOOC discussion forum contains a lot of information. Mining this text information can help teachers master learners' learning situations, topics of concern, and interest trends; analyze learners' emotional tendencies, behavior laws, and problems or trends behind these behaviors;

Zhang, Q., Sun, L. (2023). A CNN-Bi-LSTM Model for MOOC Forum Post Classification. *International Journal of Emerging Technologies in Learning (iJET)*, 18(21), pp. 89–101. https://doi.org/10.3991/ijet.v18i21.37843

Article submitted 2023-01-03. Revision uploaded 2023-07-25. Final acceptance 2023-07-25.

© 2023 by the authors of this article. Published under CC-BY.

facilitate teachers to timely adjustment of teaching contents and teaching methods; help the platform continuously optimize and improve the service; and better promote students' online learning effects. It can also improve the retention rate of students and the completion rate of courses. Therefore, mining and analyzing MOOC discussion texts has become the focus of research [2].

Classifying the text in the discussion forum is the research hotspot of text data mining and analysis in MOOCs. Text classification can help the platform optimize the navigation function of the forum and expand the learning group function, so as to improve the learners' experience of using the forum, help teachers monitor and master the dynamics of the forum in real time, quickly locate students' questions and interests, and interact effectively with students, present teachers and platform managers with posts that need help and reply, and overcome the problems of duplication of information and difficulty in screening information [3]. However, because the texts in the discussion forum are short texts, which have the characteristics of short length, less content, and colloquial expression, the text characteristics are not remarkable. At the same time, the data distribution of various categories in the MOOC discussion forum is uneven. For example, the number of posts in the homework section is obviously higher than that in the course feedback section. The unbalanced data distribution will seriously affect the classification results of the classification algorithm, which is also one of the main difficulties faced by text classification in the MOOC forum. Therefore, there are still many challenges in the research of text classification in MOOC forums that need to be further explored [4].

Aiming at the problems existing in text classification in MOOC discussion forums, this paper constructs a text classification model for MOOC forums to automatically classify the text in the discussion forum of an online learning platform, improve the automation level of the platform, and provide help and support for learners, teachers, and platform managers.

2 RELATED WORK

For a long time, text classification has been a field that researchers have paid attention to, and it is also a key application field in natural language processing. The text classification methods based on knowledge engineering cannot meet the needs of practical application because of poor flexibility, long calculation times, and application difficulties [5]. Text classification methods based on machine learning have become a research hotspot. Text classification methods based on machine learning [6] are mainly divided into two categories: traditional machine learning methods and deep learning methods.

2.1 Text classification algorithm based on traditional machine learning

Traditional machine learning algorithms include Naïve Bayes (NB), logistic regression, support vector machines (SVM), decision trees, and neutral networks (NN). Some researchers try to apply machine learning algorithms to text classification in the field of education. In [7], the author used random forest and SVM to classify the students' feedback texts and analyzed the emotional tendency of students' evaluation. Zheng et al. used four machine learning methods: NB, logistic regression, decision trees, and support vector machines to classify the posts of learners in the learning community from three dimensions of cognition, interaction and society, and analyzed the impact of different dimensions on learners' continuous learning [8].

Most of the traditional machine learning models are shallow and can effectively solve some problems under simple or specific conditions. However, when dealing with complex classification problems, there are some problems, such as low classification accuracy and weak generalization ability. In addition, traditional machine learning algorithms need to extract features manually [9]. The whole process is time-consuming and laborious, and some important features will be omitted. As to the text feature representation, the word embedding model method solves the problems of data sparsity and high dimension to a certain extent. However, it can only obtain the low-level semantic relationship of the text word vector, and it is difficult to obtain the high-level semantic dependency hidden in the text context. Therefore, the word embedding representation method has no obvious advantages over text classification algorithms based on traditional machine learning.

2.2 Text classification algorithm based on deep learning

Different from traditional machine learning methods, deep learning can automatically obtain high-level semantic features from input data and has strong expression ability. Some researchers use convolutional neural networks (CNN) for text classification [10]. Yoon Kim proposed a sentence level text classification model based on a convolutional neural network. When training word vectors, the model used CNN for pre-training to optimize feature expression and improve the classification effect [11]. Zhang Xiang et al. proposed a character level CNN model. Compared with the traditional word bag model, n-grams and word based CNN model, the classification accuracy was significantly improved [12]. However, the CNN model can capture the local organization of the text well, but it cannot capture the long-distance relationship between words in the text effectively.

In addition to CNN, recurrent neutral networks (RNNs) are also applied to text classification tasks. Aiming at the defects of the traditional English text classification algorithm, such as unclear feature items when the amount of training data is large, Liu et al. proposed a quality-related English text classification method based on RNN [13], combined with the attention mechanism, to solve the problem of label disorder, make the structure of the model more flexible, and improve the accuracy and flexibility of English text classification. A new seq2seq model based on RNN was proposed in [14], which can better capture the global potential features of sentences. The experimental results showed that the model had higher performance than the benchmark model. In theory, RNN can learn the correlation in any length sequence [15], but in fact, this is not the case. The vanishing gradient problem limits the ability of RNN to learn long-term dependencies and cannot fully obtain contextual semantic connections [16].

The long-short-term memory (LSTM) network combines short-term memory with long-term memory through gate control, solves the problem of gradient disappearance to a certain extent, and can learn long-term dependence. For example, Wei et al. proposed a learning framework based on CNN and LSTM to classify the posts in MOOC forums and determine the urgency of their needs in order to solve learners' confusion in time [17]. In order to overcome the limitation that traditional LSTM only allows the sequential transmission of information, Tai et al. proposed a tree-LSTM model [18]. Each LSTM unit can combine the information from multiple subunits so as to obtain richer semantic information. However, the model's performance depends on the size of the data set and the selection of the alphabet. Both RNN and LSTM can only predict the output of the next time based on the information of the previous time. However, in some problems, the context should be fully considered. Not only the previous text but also the subsequent text should be taken into account. Bi-directional LSTM (Bi-LSTM) can solve this problem [19]. A classification model based on Bi-LSTM was proposed in [20]. Bi-LSTM was used to extract the context representation before and after the phase, and a softmax classifier was used to classify the processed context information, which effectively improved the accuracy of classification.

This paper aims to combine CNN and Bi-LSTM to build a model for MOOC forum post classification. CNN is used to extract local key semantic information from text. Bi-LSTM is used to extract the context-semantic relationship of text from a global perspective. Thus, the local semantic features and the global context features of the text are fully considered to achieve higher text classification accuracy.

3 CNN-BI-LSTM TEXT CLASSIFICATION MODEL

This paper proposes a text classification model integrating CNN and Bi-LSTM to classify the posts in MOOC forum, as shown in Figure 1.

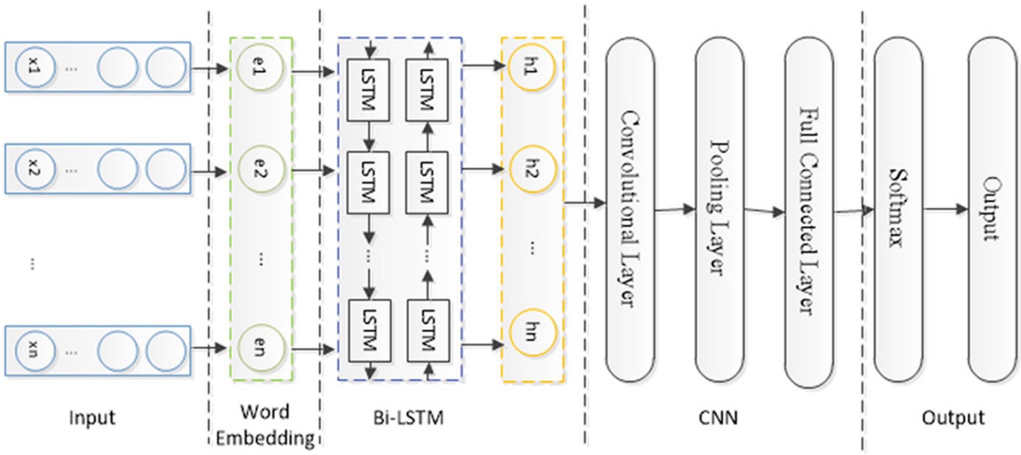


Fig. 1. The framework of CNN-Bi-LSTM text classification model

The discussion posts are preprocessed by word segmentation, removing stop words, and then input into the classification model. The word embedding layer receives text data from the input layer and converts the text data into word vectors to obtain a shallow semantic representation of the text. The Bi-LSTM layer is responsible for extracting the context information of the text vectors to obtain the deep semantic dependency hidden in the text context. In the convolution layer, the CNN network is used to extract the local features of the text, and multiple maximum features are extracted through the pooling layer to enter the full connected layer. Finally, the softmax function is used to normalize the feature vectors to obtain the final output of the classification prediction value.

3.1 Bi-LSTM layer

Traditional text classification models mostly use a single convolutional neural network. The text features, represented by a two-dimensional matrix, are input into the model. However, the semantic relationship of sentence context is also crucial to the understanding of sentences. The features extracted using CNN alone do not contain contextual semantic information, which will affect the final classification accuracy. Therefore, a Bi-LSTM layer is added to the classification model to extract the context information of the text. The output of this layer is obtained by splicing the forward LSTM output and the backward LSTM output [21], as illustrated in Figure 2.

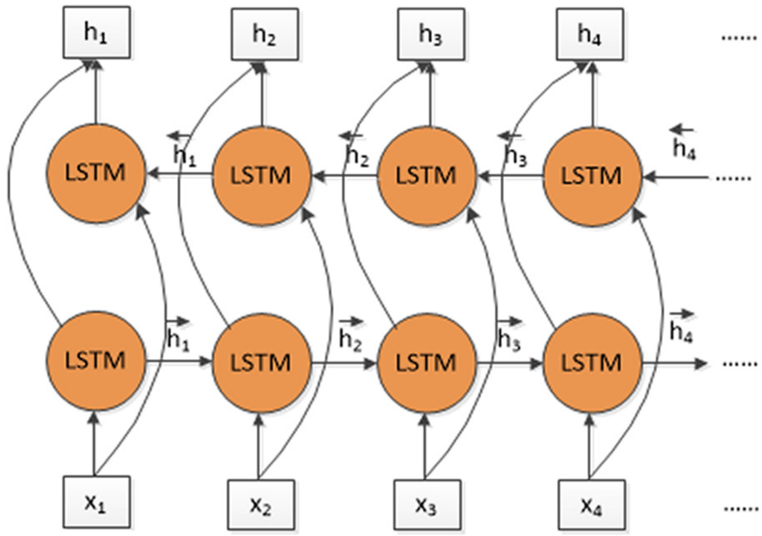


Fig. 2. Bi-LSTM structure diagram

The model uses bi-directional semantic dependency extracted from Bi-LSTM layer to reflect the core features of text through weighting matrix, hoping to make up for the shortcomings of single CNN text classification model and improve the accuracy of classification.

The dropout method is used in the Bi-LSTM layer to prevent the model from over-fitting. In the training phase, dropout randomly discards some nodes according to probability p, but retains their weights. These weights are restored in the next training, and then randomly discard some nodes again. Repeat this process, which is equivalent to training different local networks each time, for only the local network is trained each time, the training time is effectively reduced. On the other hand, in the test phase, all abandoned nodes are recovered, and all local networks are combined to improve the generalization capability of the model. The calculation method for dropout is shown in (1).

$$h = \delta f(Wx), \delta \sim Bernoulli(p) \tag{1}$$

Where, x is the input of the layer, W is the weight value, f is the activation function, δ is the dropout mask, and the probability is p that each element in δ is 1.

3.2 CNN layer

In the proposed model, CNN is used to extract the local feature vectors of the text. The hidden state sequence of the text obtained through Bi-LSTM is represented by H, $H = (h_1, h_2, ..., h_n)$, where n represents the length of the text. The vectors $h_i, h_{i+1}, ..., h_{i+k-1}$ are spliced, where h_i represents the ith single word vector, and k represents the size of the convolution kernel, then the text feature of the ith window is extracted using (2).

$$\begin{cases} Y_i = g(x_i \times W + b) \\ x_i = \oplus(h_{i:i+k-1}) = \left[h_i, h_{i+1}, \ldots, h_{i+k-1}\right] \end{cases} \tag{2}$$

Where, Y_i represents the ith vector of convolution output, x_i represents the splicing input vector of the ith window; ⊕ represents the splicing operation, b is the offset vector and W is the weight matrix.

The convolution kernel slides from the first word to the last word on the input text sequence to extract the text features. Convolution operation is defined as (3).

$$Conv_{U,b}^{k,m} \tag{3}$$

Where, k is the size of the convolution kernel, m represents the length of the sliding window. The text feature extraction process in the convolution layer can be expressed as (4). Then, max pooling is performed on $Y_{1:m}$ to extract the most important features in the text sequence.

$$Y_{1:m} = Conv_{U,b}^{k,m}(S_{1:n}) \tag{4}$$

4 EXPERIMENTS

To evaluate the effectiveness of the proposed CNN-Bi-LSTM classification model, we test our method and compared with other benchmark model.

4.1 Dataset

The dataset was collected from the discussion forum of 10 popular courses on the MOOC platform of Chinese universities. The data set contains 19285 posts, which are divided into four categories (see Table 1), including course content, course logistics, evaluation suggestions, and others. Each category of data was then divided into training and test sets according to the data division ratio of 8:2. Table 2 shows the distribution of the four categories of the discussion text.

Table 1. Text category division

Category Code	Category Name	Description	Example
C	Course content	Discussion posts directly related to the course content, mainly including questions, answers, comments and views related to the course content.	Could you tell me the difference between primary key, primary code, and keyword?
L	Course logistics	Discussion posts related to course logistics mainly include questions and explanations related to course schedule, course materials, exams, certificates and platforms.	How can I obtain a certificate? When will we take the exam?
E	Evaluation suggestions	Comments and suggestions on MOOC platform, courses and teachers.	The teacher explained thoroughly, very well!
N	Others	Posts unrelated to the above categories.	Go China!

Table 2. Distribution of various categories of posts

Categories	Number of Posts
Course content (C)	6815
Course logistics (L)	5575
Evaluation suggestions (E)	3998
Others (N)	2897

4.2 Experiment setting

In order to verify the effect of the CNN-Bi-LSTM text classification model proposed in this paper, the experiment was conducted to compare it with the five baseline models, which are LightGBM, NB, LSTM, Bi LSTM, and CNN. The experiment platform is Google TensorFlow, the programming language is Python, and the development tool is PyCharm. The parameters of the model have a significant impact on the experimental results, and in our experiments, the main parameter settings of each classification model are shown in Table 3.

Table 3. Main parameter settings of the experiment

Parameter	Instructions	Value
dropout	the loss rate	0.5
num_classes	the number of categories	4
num_epochs	the number of iterations	20
batch_size	the size of mini-batch	128
learning_rate	the learning rate	1e-3
filter_sizes	the size of convolutional kernel	(2, 3, 4)
num_filters	the number of convolutional kernel	256
hidden_size	the number of Bi-LSTM hidden cell	128
embedding_size	the dimension of the word vector	300

4.3 Evaluation criterion

For each category, the classification results of the model include four types, as shown in Table 4. Each category treats itself as "positive" and all other categories as "negative."

Table 4. Confusion matrix

	Number of posts predicted by the classification model to be positive	Number of posts predicted by the classification model to be negative
Number of posts that are actually positive	TP	FN
Number of posts that are actually negative	FP	TN

In this paper, accuracy, precision, recall, and F1 score are used as evaluation indicators of text classification model.

1. Accuracy

Accuracy is used to describe the proportion of all correctly predicted samples in the total number of samples. The calculation method is shown in (5).

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

2. Precision

Precision represents the proportion of the predicted actual positive samples in the predicted positive samples. The calculation method is shown in (6).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{6}$$

3. Recall

Recall represents the proportion of positive samples that are correctly predicted. The calculation method is shown in (7).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{7}$$

4. F1 score

F1 score is an indicator that comprehensively considers the accuracy rate and recall rate, which is calculated by (8).

$$\text{F1} = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

4.4 Results and analysis

The proposed CNN-Bi-LSTM text classification model was compared with five benchmark models on the preprocessed dataset to evaluate the effectiveness of the proposed model, where LightGBM (light gradient boosting machine) and NB are traditional machine learning models, and LSTM, Bi-LSTM and CNN are deep learning models. The comparison results are shown in Table 5.

Table 5. Evaluation of each classification model

Model	Evaluation Criterion				Accuracy
	Category	Precision	Recall	F1 Score	
LightGBM	C	0.71	0.96	0.81	0.83
	L	0.89	0.63	0.73	
	E	0.93	0.79	0.85	
	N	0.79	0.90	0.94	
NB	C	0.81	0.91	0.86	0.87
	L	0.87	0.75	0.81	
	E	0.89	0.89	0.89	
	N	0.98	0.89	0.93	
LSTM	C	0.89	0.88	0.88	0.89
	L	0.91	0.93	0.92	
	E	0.86	0.82	0.84	
	N	0.92	0.96	0.94	
Bi-LSTM	C	0.89	0.93	0.91	0.91
	L	0.89	0.95	0.92	
	E	0.95	0.83	0.89	
	N	0.97	0.93	0.96	

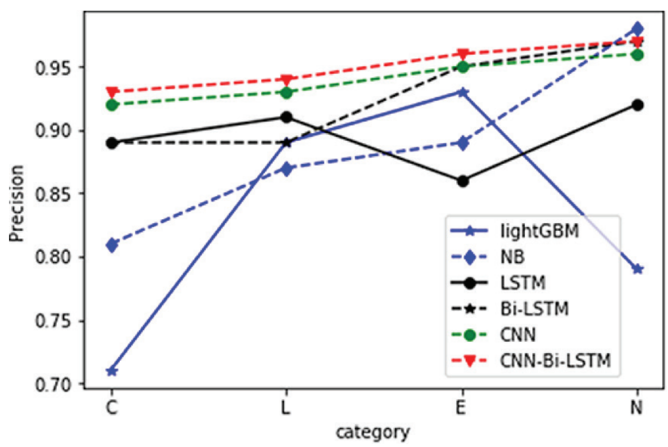
(Continued)

Table 5. Evaluation of each classification model *(Continued)*

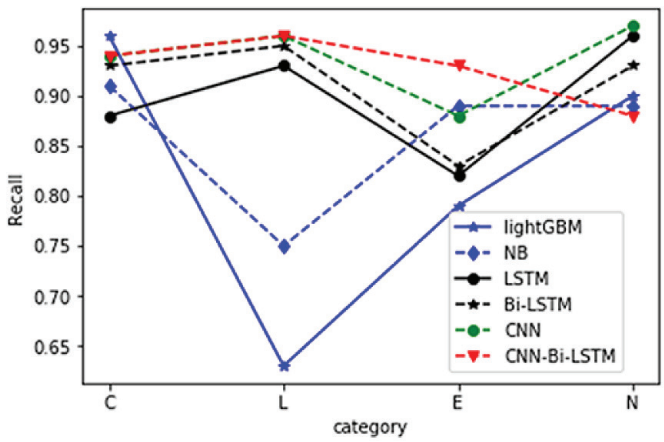
Model	Evaluation Criterion				Accuracy
	Category	Precision	Recall	F1 Score	
CNN	C	0.92	0.94	0.93	0.93
	L	0.93	0.96	0.94	
	E	0.95	0.88	0.92	
	N	0.96	0.97	0.96	
CNN-Bi-LSTM	C	0.93	0.94	0.93	0.95
	L	0.94	0.96	0.95	
	E	0.96	0.93	0.94	
	N	0.97	0.98	0.97	

The results show that the classification models (LSTM, Bi-LSTM, CNN, CNN-Bi-LSTM) based on deep learning are superior to the traditional machine learning classification models (LightGBM, NB). Compared with the traditional machine learning algorithms, the deep learning algorithms have stronger parallel processing and learning abilities and are not vulnerable to noise in the data. Although deep learning algorithms need more parameters and consume more model training time, their classification accuracy and stability are higher than those of traditional machine learning algorithms. According to the comprehensive evaluation indicators, the proposed CNN-Bi-LSTM model achieves better classification results than other baseline models, and the classification accuracy rate reaches 95%. Whether the total classification accuracy of the model or the classification accuracy of a single category post is better than the other five machine learning algorithms. The F1 score of each category post reaches 93%, 95%, 94%, and 97%, respectively.

Figure 3 shows the classification precision, recall and F1 score of the six classification algorithms on posts of different categories.

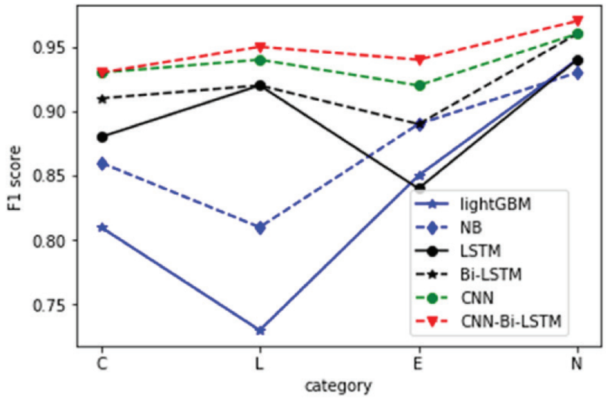


a) Precision



b) Recall

Fig. 3. *(Continued)*



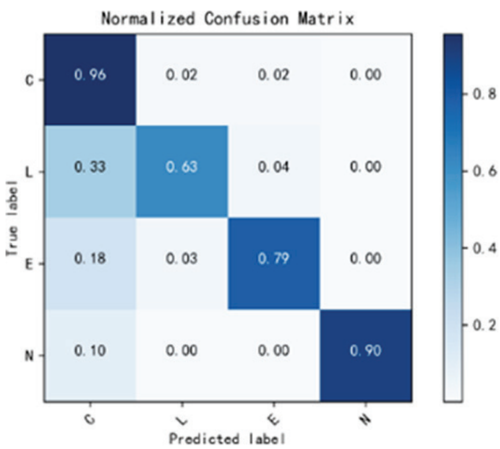
c) F1 score

Fig. 3. Precision, recall and F1 score of different models

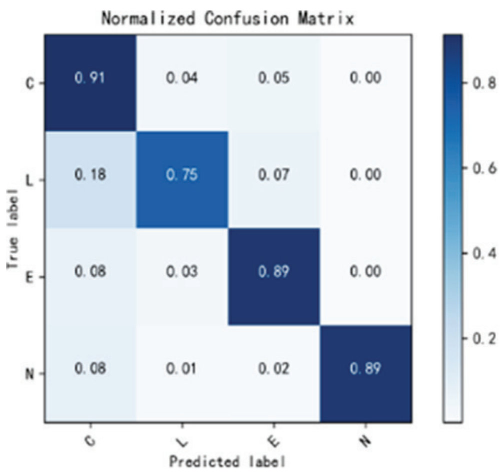
It can be seen from Figure 3 that the precision of the CNN-Bi-LSTM model is higher than that of other models. According to the smoothness of the curve, the LightGBM model has the worst classification stability, and the classification precision of CNN and CNN-Bi-LSTM on different categories of posts remains relatively stable. Compared with the other three categories, the classification precision of the LSTM model on posts in category E (evaluation suggestions) decreases because this type of post contains more words expressing emotions. The Bi-LSTM model solves this problem by extracting semantic dependencies between contexts and greatly improving the precision of the evaluation suggestion posts.

The traditional machine learning algorithms LightGBM and NB have similar trends in recall across different categories. On the whole, the CNN-Bi-LSTM model performs best. From the curve smoothness, the F1 score of each model is more stable than the recall. The F1 score in category N is the highest. Although the sample in this category is the least, the post-text features are obvious. The F1 scores of the CNN-Bi-LSTM model for classifying four categories of posts are higher than those of other benchmark models.

Figure 4 shows the normalized confusion matrix of each classification model, which more intuitively shows the classification effect of each model on different categories of posts.



a) LightGBM



b) NB

Fig. 4. *(Continued)*

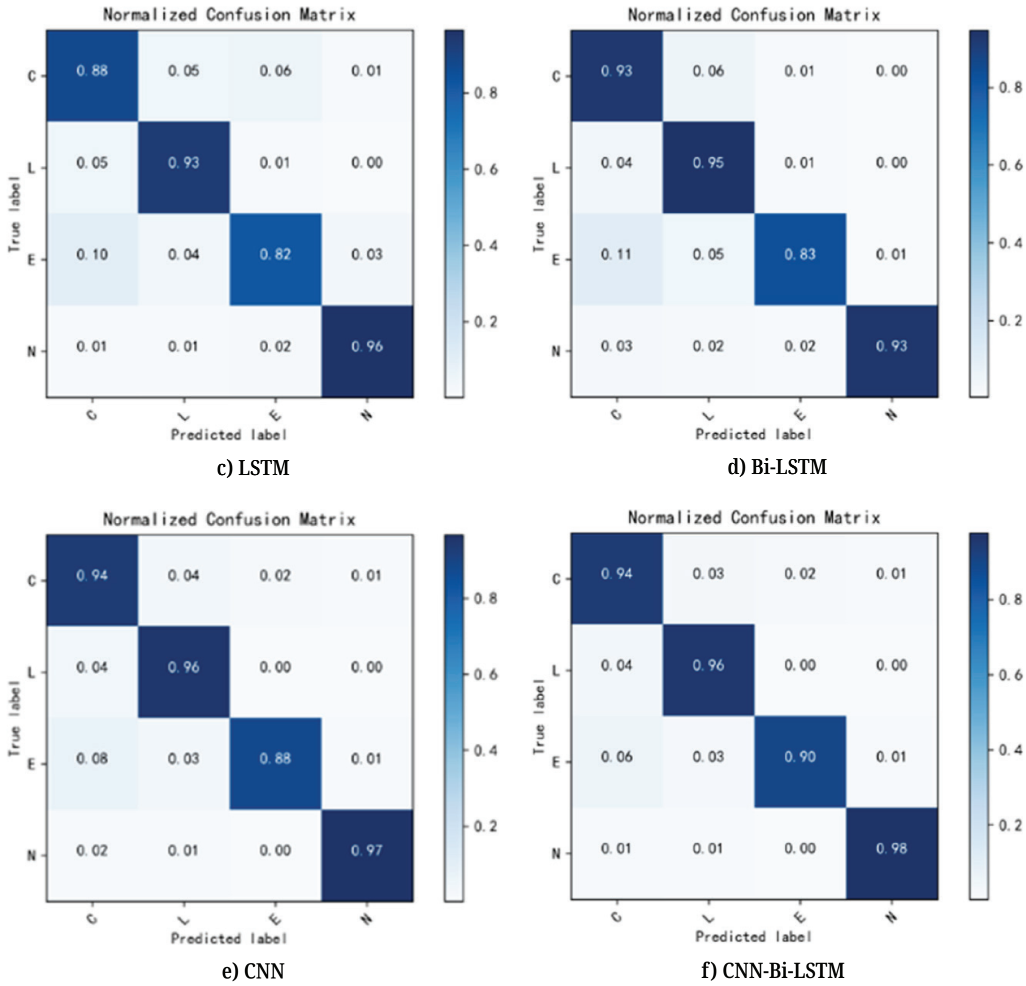


Fig. 4. Confusion matrix diagram

Compared with the traditional machine learning models, the deep learning models have a more stable classification effect on different categories of posts. It can be seen from Figure 4(c) that when the LSTM model classifies the course content-related posts (C), 6% of them are classified as evaluation suggestion posts (E), and 5% are classified as course logistics posts (L). The normalized confusion matrix coefficient of this type of post is 88%, and the correct classification probability of the Bi-LSTM model on the course logistics posts has reached 93%, which is 5% higher than that of LSTM. The CNN-Bi-LSTM model proposed in this paper performs best, and the normalized confusion matrix coefficients of the four categories of posts reach more than 90%.

5 CONCLUSION

This paper proposes a text classification model based on CNN and Bi-LSTM. This model combines the advantages of CNN and Bi-LSTM. Bi-LSTM is used to obtain global context information, and CNN is used to extract local key semantic information. Experiments were carried out on a data set containing 19285 discussion posts from the MOOC platform of the University of China and compared with five benchmark models. The results show that the overall classification accuracy of the proposed CNN-Bi-LSTM model is 95%, which is 12%, 8%, 6%, 4%, and 2% higher than LightGBM, NB, LSTM, Bi-LSTM, and CNN, respectively. This method has high application value in automatic text classification, teaching supervision and intervention, platform assistance, and support in MOOCs. However, the model still has some shortcomings. The experimental results show that the accuracy of the model is relatively low in distinguishing between evaluation suggestions, course content, and course logistics posts. In our future research, we will try to design a more complex and efficient text classification model to optimize the classification effect and improve the accuracy of text classification.

6 ACKNOWLEDGMENT

This work is partially supported by the 14th Five-Year Plan project for Educational Science in Henan Province (No. 2021YB0124). The authors also acknowledge the comments and suggestions of the reviewers, which proved to be helpful in improving the presentation.

7 REFERENCES

[1] X. Gu, "A research based on MOOC course comment for learning behavior analysis," Cenral China Normal University, Wuhan (in Chinese), 2018.

[2] H. X. Wang, Z. M. Yan, X. Y. Chen, and M. R. Zhang, "Research on topic mining and emotion analysis for MOOCs course review," *Journal of Open Learning*, vol. 26, no. 4, pp. 16–23, 2021.

[3] J. Chen, J. Feng, X. Sun, and Y. Liu, "Co-training semi-supervised deep learning for sentiment classification of MOOC forum posts," *Symmetry*, vol. 12, no. 1, pp. 8–20, 2019. https://doi.org/10.3390/sym12010008

[4] R. Alnusyan, R. Almotairi, S. Almufadhi, A. A. Al-Shargabi, and J. F. Alshobaili, "Hybrid approach for user reviews' text analysis and visualization: A case study of amazon user reviews," *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 16, no. 8, pp. 79–93, 2022. https://doi.org/10.3991/ijim.v16i08.30169

[5] G. C. Liu, "MOOC forum discussion threads classification research based on machine learning," Dalian University of Technology, Dalian (in Chinese), 2018.

[6] L. Almazaydeh, M. Abuhelaleh, A. Al Tawil, and K. Elleithy, "Clinical text classification with word representation features and machine learning algorithms," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 19, no. 4, pp. 65–76, 2023. https://doi.org/10.3991/ijoe.v19i04.36099

[7] L. Feng, G. Liu, S. Luo, and S. Liu, "A transferable framework: Classification and visualization of MOOC discussion threads," in *Proceeding of the International Conference on Neural Information Processing*, Springer, Cham, 2017, pp. 377–384. https://doi.org/10.1007/978-3-319-70093-9_39

[8] W. Xing and G. Fei, "Exploring the relationship between online discourse and commitment in twitter professional learning communities," *Computers & Education*, vol. 126, pp. 388–398, 2018. https://doi.org/10.1016/j.compedu.2018.08.010

[9] Y. F. Zheng, Z. H. Gao, J. Shen, and X. S. Zhai, "Optimising automatic text classification approach in adaptive online collaborative discussion—A perspective of attention mechanism-based Bi-LSTM," *IEEE Transactions on Learning Technologies*, 2022. https://doi.org/10.1109/TLT.2022.3192116

[10] M. Niu, "Classification of learning sentiments of college students based on topic discussion texts of online learning platforms," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 24, pp. 42–56, 2022. https://doi.org/10.3991/ijet.v17i24.35951

[11] Y. Kim, "Convolutional neural networks for sentence classification," *Association for the Computation Linguistics*, vol. 15, no. 6, pp. 1746–1751, 2014.

[12] X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," in *Proceeding of the International Conference on Neural Information Processing Systems*, Mit Press, 2015, pp. 649–657.

[13] C. Liu and X. Wang, "Quality-related english text classification based on recurrent neural network," *Journal of Visual Communication and Image Representation*, vol. 71, p. 102724, 2020. https://doi.org/10.1016/j.jvcir.2019.102724

[14] M. Jang, S. Seo, and P. Kang, "Recurrent neural network-based semantic variational autoencoder for sequence-to-sequence learning," *Information Sciences*, vol. 490, pp. 59–73, 2019. https://doi.org/10.1016/j.ins.2019.03.066

[15] B. Jiang, S. Wu, C. Yin, and H. Zhang, "Knowledge tracing within single programming practice using problem-solving process data," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, pp. 822–832, 2020. https://doi.org/10.1109/TLT.2020.3032980

[16] Z. Peng, S. Wei, J. Tian, Z. Qi, *et al.*, "Attention-based bidirectional long short-term memory networks for relation classification," in *54th Annual Meeting of the Association for Computational Linguistics*, Berlin, Germany, 2016, pp. 207–212.

[17] X. Wei, H. Lin, L. Yang, *et al.*, "A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification," *Information*, vol. 8, no. 3, p. 92, 2017. https://doi.org/10.3390/info8030092

[18] K. S. Tai, R. Socher, and C. D. Manning, "Improved semantic representations from tree-structured long short-term memory networks," in *53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, Beijing, China, 2015, pp. 1556–1566. https://doi.org/10.3115/v1/P15-1150

[19] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5–6, pp. 602–610, 2005. https://doi.org/10.1016/j.neunet.2005.06.042

[20] G. Liu and J. Guo, "Bidirectional LSTM with attention mechanism and convolutional layer for text classification," *Neurocomputing*, vol. 337, pp. 325–338, 2019. https://doi.org/10.1016/j.neucom.2019.01.078

[21] Dilek Hakkani-Tür, G. Tur, A. Celikyilmaz, *et al.*, "Multi-domain joint semantic frame parsing using Bi-directional RNN-LSTM," in *17th Annual Meeting of the International Speech Communication Association*, Interspeech, 2016, pp. 715–719. https://doi.org/10.21437/Interspeech.2016-402

8 AUTHORS

Qiaorong Zhang is a Professor at College of Computer and Information Engineering, Henan University of Economics and Law, China. Her research interests include learning analysis technology, education data mining, and AI in education.

Lin Sun is a PhD student at School of Economics and Management, China University of Petroleum (Beijing), China. Her research interests include intelligent decision making and optimization, industrial internet platform operation and management, and manufacturing resource allocation.