IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 23 February 2024, accepted 1 June 2024, date of publication 11 June 2024, date of current version 20 June 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3412649

RESEARCH ARTICLE

Chinese Multilabel Short Text Classification Method Based on GAN and Pinyin Embedding

JINPENG BAI [1] **AND XINFU LI** [1,2]
[1]School of Cyber Security and Computer, Hebei University, Baoding 071000, China
[2]Hebei Machine Vision Engineering Research Center, Baoding 071000, China
Corresponding author: Xinfu Li (mc_lxf@126.com)

ABSTRACT With the development of the Chinese Internet, a large amount of Chinese short text data has been generated. The multilabel classification of Chinese short texts enables more effective management and analysis. However, due to the sparsity of Chinese short text features, and the fact that commonly used multilabel classification models are primarily designed and developed in English, traditional sampling methods can easily lead to poor classification results. In response to these challenges, we propose a Chinese multilabel short text classification method based on GAN and enhanced with pinyin. Firstly, we utilize BERT, augmented by pinyin embedding, as a method for text vector representation to enrich text information. Secondly, multiple hidden layers of BERT are integrated with the generators of the GAN model to comprehensively learn the feature distribution. Finally, the improved sampling method is used to help the model learn better. Experimental results show that the method proposed in this article performs better in processing Chinese multilabel short text classification tasks.

INDEX TERMS GAN, natural language processing, text classification.

I. INTRODUCTION

With the development of China's mobile Internet, vast amounts of text data are generated daily from platforms such as search engines, social networks, and news websites. Chinese short text data continues to increase, covering various fields such as news topics, personal emotions, and email messages. In order to make reasonable use of information, how to classify Chinese online short text has become an important research field.

As one of the essential tasks in natural language processing, short text classification is widely used in spam filtering, personalized recommendations, and the implementation of intelligent question-and-answer robots. In short text classification, the task of classifying a text into different labels is called multilabel short text classification. For example, "What will P2P financial management look like after rectifying the industry?" This news is categorized under both technology and finance. As the number of Chinese short texts continues to increase, accurate multilabel classification of Chinese short text is essential. Therefore, the Chinese multilabel short text classification task has gradually attracted people's attention.

As a branch of multilabel text classification, the Chinese multilabel short text classification has similar solution ideas to multilabel classification. For example, traditional multilabel classification methods include problem transformation methods (such as CC [1]) and adaptive methods (such as ML-KNN [2]). The main idea is to convert the multilabel classification problem into a single-label classification problem or convert a single-label classifier into a multilabel classifier according to the task. Traditional classification methods can perform most Chinese multilabel short text classification tasks but rely too much on feature engineering. Compared with traditional classification methods, deep learning models can obtain better classification results without feature engineering, such as convolutional neural network (CNN) [3], recurrent neural network (RNN) [4], attention mechanism [5] etc. Generative Adversarial Networks (GAN) have gained widespread attention from researchers as they effectively mitigate the impact of imbalanced sample distribution on classification accuracy. However, in text classification tasks,

The associate editor coordinating the review of this manuscript and approving it for publication was Kashif Saleem.

when training GAN, traditional sampling methods cause the model to have difficulty with convergence.

Additionally, Chinese online short text exhibits characteristics such as sparsity, real-time nature, and irregular grammar. In order to fully utilize the Chinese text information, many Chinese representation methods have been proposed, such as word embedding model CWE [6] using radical segmentation; Cw2Vec [7] using Chinese sub-word information; using char-acters, radicals, Wubi code, and pinyin code model Moto [8]; Chinese dynamic pre-training language model ERNIE [9] combined with contextual information. The above models show that adding free radicals can help solve the sparsity problem of Chinese short text data [10].

In summary, although people have achieved good results in multilabel classification and Chinese short text semantic feature learning, only a few models can organically combine the two and improve the accuracy of Chinese multilabel short text classification. During model training, there is a problem of difficulty in model convergence due to traditional sampling. With the rapid development of the Chinese mobile Internet, there is an urgent need to develop models suitable for Chinese multilabel short text classification. Therefore, this paper builds a Chinese multilabel classification model that combines pinyin embedding and GAN. The main tasks are as follows:

We proposed a Chinese multilabel short text classification method. We incorporate the pinyin information of Chinese characters into the pre-training model to generate text representations containing pinyin embeddings to enrich text information; connect multiple hidden layers of BERT to the generators to increase the probability that the distribution of negative samples is similar to that of positive samples; we also propose an improved sampling method which can reduce the impact of positive and negative samples being too similar, causing difficulty in model convergence.

The proposed method was tested on a data set of 520 thousand Toutiao headlines and 570 thousand questions and achieved good results, verifying the model's effectiveness for the Chinese multilabel short text classification task.

II. RELATED WORK

A. MULTILABEL CLASSIFICATION

The task of multilabel text classification aims to predict multiple category labels for a given text, where the labels are not mutually exclusive. As deep learning models have made significant achievements in text classification, more and more researchers are using deep learning models to solve multilabel text classification problems. Ma et al. [11] combined BiGRU-based text representation with category structure-based graph embedding, using category label-based word embedding for hierarchical text representation, and multi-layer perceptron for multilabel classification; Liu et al. [12] based on CNN, introduced dynamic pooling and hidden bottleneck layers to accomplish multilabel text classification; Chen et al. [13] proposed a classification model based on label embedding and capsule neural network to achieve multilabel classification of legal texts; Fei et al. [14] learned the hidden emotion of the text is predicted through BiGRU to achieve emotion distribution without external knowledge and complete multilabel emotion classification; Yan et al. [15] combined CNN and capsule network to fulfill the multilabel text classification task.

B. CHINESE TEXT CLASSIFICATION

The above-mentioned deep learning models have achieved good results in multilabel classification. However, most of them are designed based on English. Chinese and English are different, resulting in text structure, pinyin tones, and other information needing to be fully utilized, resulting in Chinese text classification tasks. The effect could be better [16].

With the widespread application of Chinese text in Internet systems, more and more classification models designed for Chinese have been proposed. Sun et al. [10] proposed to use 6-granularity features such as word-jieba, word-jieba-radical, and word-ngram-radical to build a Chinese short text classification model and verify the effectiveness of the model; Liu [17] used GCN with BERT as the framework. Abstract meanings are used to construct event graphs, and candidate labels are obtained with the help of multi-hop storage networks, to complete the Chinese multilabel text classification task; Wang and Guo [18] extracted Chinese semantics through XLNet, BiLSTM and HA embedding, classifying solve the problem of multilabel classification of poverty governance texts; Tang et al. [8] analyzed Chinese character information and integrated the pinyin code, five-stroke structure, Chinese character roots, and Chinese characters themselves through the attention mechanism, and through complete learning Semantic information to improve multi-classification accuracy of Chinese short text.

C. GAN MODEL

With the increased text and the increasing number of unlabeled data and small sample labels, researchers are gradually paying attention to semi-supervised learning methods. GAN was first applied to machine vision. The main idea is to use a generator to construct a negative sample set and combine it with positive samples to train the discriminator. As researchers pay more and more attention to unlabeled samples, GAN models are gradually applied to natural language processing. For example, Croce et al. [19] integrated the GAN model into BERT and completed the text classification task through two-stage training; Auti et al. [20] used BioBERT to capture text specificity and combine it with improved GAN to classify medical text; Shehnepoor et al. [21] used review text and rating scores to improve GAN in the process of generation and detection. They proposed a ScoreGan model based on data enhancement to complete a fraud review; Li et al. [22] classified Chinese government cable texts by combining BERT with embedded pointers and the GAN model.

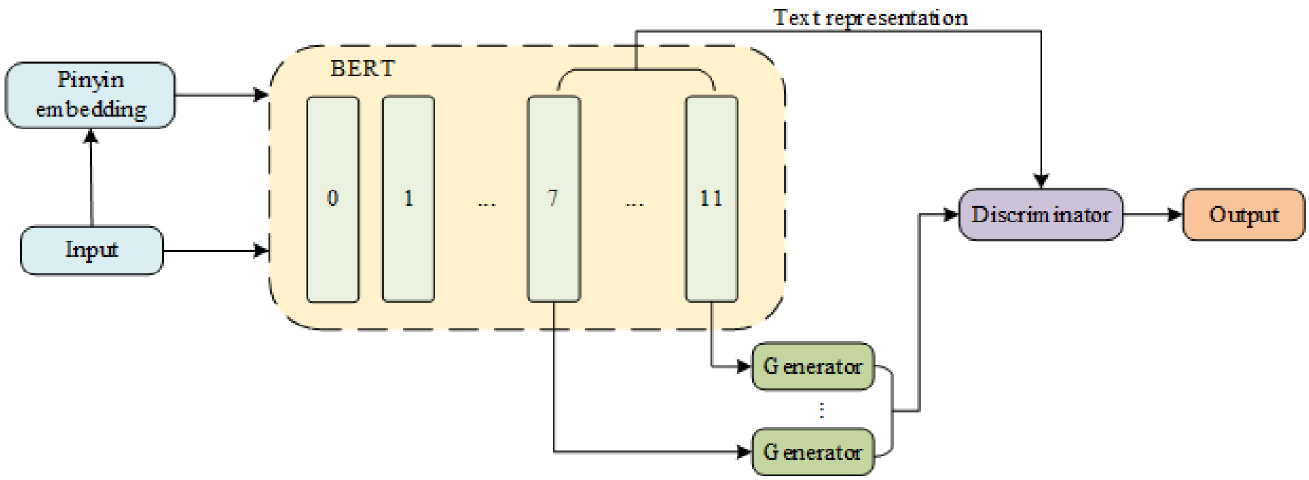


FIGURE 1. Model framework. It is mainly composed of two parts: text representation and classification.

In summary, the existing multilabel classification models can effectively deal with English and Chinese long texts. However, due to the characteristics of sparse features and the broad classification of Chinese short text, the classification results of Chinese multilabel short text could be better. The GAN model can better cope with the problem of unbalanced sample distribution. However, there is a problem during training: the model needs help in convergence due to too many positive and negative samples.

III. PROPOSED METHOD

A. FRAMEWORK

Figure 1 shows the overall structure of the model. In the text representation part, first, the embeddings of text characters and pinyin serve as the model's input; then, they are mapped to the d-dimensional fusion embedding through the mapping layer; finally, the fusion embedding replaces the character embedding in the BERT model during training. In the classification part, first, the generators are connected to the last five layers of the BERT model. Negative samples are obtained through improved sampling methods, and finally, the discriminator is trained using negative and positive samples.

B. TEXT REPRESENTATION

1) PINYIN EMBEDDING

For Chinese characters, pinyin is highly related to semantics [8]. A Chinese character may have multiple pronunciations with different semantics. Therefore, this article attempts to weight the input of the BERT model through tones, improve the BERT model, and embed the pinyin of the text in the semantic learning process. The specific composition is shown in Figure 2:

An open source pinyin package generates a pinyin sequence containing tones, and multi-phonetic words are standardized through the pinyin sequence. This article fixes the length of the pinyin sequence of a single text to 8, places the tone marker after each pinyin syllable. The resulting pinyin sequence is then input to a CNN model with a width of 2, and the pinyin embedding is obtained through the maximum pooling layer.

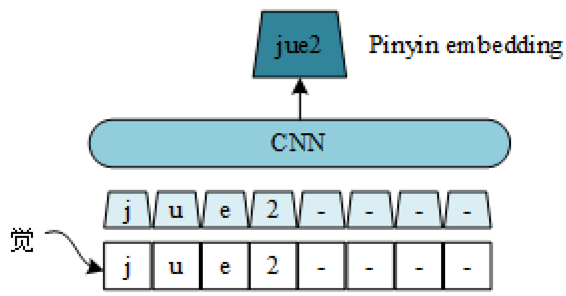


FIGURE 2. Pinyin embedding.

2) FUSION EMBEDDING

In fusion embedding, character-level token embedding and pinyin embedding are concatenated to form a 2-D dimensional vector. The fusion vector is obtained through Equation (1) in the fully connected layer, as shown in Figure 3. The fusion vector is input into BERT instead of token embedding for training.

$$V_F = W_F V_C + b \tag{1}$$

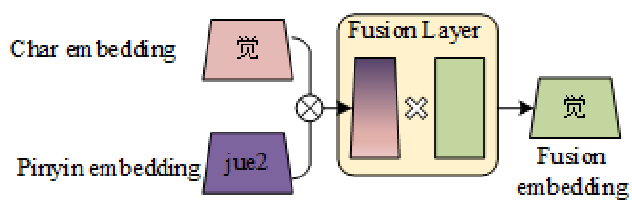


FIGURE 3. Fusion embedding. The 2-D dimension vector will be mapped to the 1-D dimension.

Among them, W_F is the weight matrix, V_C is the concatenate of token embedding and pinyin embedding, and b is the offset.

C. CLASSIFICATION

Most GAN models usually adopt a learning method of alternating training of the generator and discriminator. The model's performance depends to a certain extent on how well the generative model learns the sample distribution, especially when there are few training samples. To improve the probability that the distribution of negative samples is similar to actual samples, multiple hidden layers of the BERT model are used to connect to the generator, and different hidden layers contain different language information to learn the sample distribution.

Since the same sentence will have different representations in different hidden layers, the generators will perform a mask operation on words at the same random position in the sentence. The generator calculates the probability of each word in the word list through Equation (2) for the mask mark position based on the context information. It uses the word with the highest probability as the word being asked. If the word with the highest probability is not the original asked word, the sample is regarded as a fake sample; otherwise, it is a positive sample.

$$p_g(x_t) = soft\max(e(x_t)^T g(\tilde{x})) \tag{2}$$

Among them, $e(x)$ is the vector representation of the word x, and $g(x)$ is the output of the hidden layer. By combining the fake samples corresponding to the same actual sample of each generator through Equation (3), the final representation f of the fake sample is obtained:

$$f_g(x_t) = avg(g_f(x_t)_i), i \in [1, 5] \tag{3}$$

Among them, $g_f(x_t)_i$ is the vector representation of the i-th fake sample.

In traditional models, the collection of negative samples occurs after the training phase is completed. When the negative samples are too similar to the positive samples, and since each fake sample has the same probability of being collected, it can easily cause difficulty in model convergence [23]. When the generator proposed in this article builds a negative sample set, it will score the fake samples and distinguish the difficulty based on the scores. In the early stages of model training, a larger gap between the negative sample and the actual data facilitates the discriminator's learning of text features. As the model is continuously trained, the fake samples generated by the generator become more and more similar to the actual samples, and the quality of the negative sample set obtained from this also improves, and the discriminator will receive better training. This enhances the discriminator's ability to distinguish, thus aiding in its training from easy to difficult tasks.

Therefore, a function needs to be constructed so that the sampling probability decreases as the similarity between the fake sample and the positive sample increases. The generator contains a fully connected layer using the sigmoid function, which evaluates the samples while generating fake samples, as shown in Equation (4):

$$G(f) = \sigma(W_g f + b) \tag{4}$$

Among them, W_g is the weight matrix, and f is the vector representation of the fake sample. The score of the fake sample is calculated through the sigmoid function. The closer the score is to 1, the more similar the fake sample is to the actual sample. By evaluating the fake samples, a negative sample set is obtained S_g, as shown in Equation (5):

$$S_g = \{l, l \in S_a\} \tag{5}$$

S_a the set of fake samples generated by the generator l is the sample in the set of fake samples. During training, in an epoch loop, the generator in the current state will score the samples in the set of fake samples, recalculate the probability of being extracted, and select fake samples from easy to challenging ones, ultimately forming a negative sample set. The loss function of the generator is shown in Equation (6):

$$\begin{aligned} L_g(G(f), y_g) = \sum_{i=0}^{N} (1 - y_g^i)(-\log(1 - G(f)^i)) \\ + y_g^i(-\log(G(f)^i)) \end{aligned} \tag{6}$$

where y_g is the multi-hot vector representation of the label corresponding to the text.

In the Chinese multilabel short text classification task, since the text data are discrete, it is difficult for the GAN model to update the generator's parameters using gradients from the discriminator. Therefore, we only use the actual sample loss value to update the generator parameters while adding actual data to the positive sample collection.

When S_g vectorization, it is used as the discriminator input with the positive sample set S_p. Among them, the positive sample is a vector represented by the mean value of the output of the last five hidden layers. After the input sample vector is convolutionally pooled, the significant feature $C_{\max}$ is obtained, and the score D of the sample vector for different labels is calculated through Equation (7):

$$D(e) = \sigma(MC_{\max}) \tag{7}$$

Among them, M is the label embedding. The expression of the discriminator is shown in Equation (8):

$$\begin{aligned} L_d(D(e), y_d) = \sum_{i=0}^{N} (1 - y_d^i)(-\log(1 - D(e)^i)) \\ + y_d^i(-\log(D(e)^i)) \end{aligned} \tag{8}$$

The overall loss of the model is obtained by summing the losses of the generator and the discriminator, as shown in Equation (9).

$$L = L_g + L_d \tag{9}$$

During training, first, fake samples are generated by multiple generators; second, a set of negative samples is obtained through sampling; third, positive samples and negative samples are input into the discriminator for training, and the model loss L is calculated; finally, the generator is updated according to L, discriminator and hidden layer parameters.

IV. EXPERIMENT

A. DATASETS

Few data sets can be directly used for Chinese multilabel short text classification. Therefore, a Chinese multilabel short text data set is constructed by extracting data from existing Chinese data sets for evaluating the model.

When extracting from a large dataset, the first step is to filter out samples that contain multiple labels. Then, all samples with these labels are extracted to form a multilabel classification dataset. Below is an introduction to the extracted dataset.

As a large Chinese media platform, Toutiao has a vast number of Chinese articles, and each article has been manually classified. This work extracts the Toutiao headline data set (https://github.com/aceimnorstuvwxz/toutiao-multilevel-text-classfication-dataset)and builds a title classification data set. The title data are divided into 28 classes and contains 520,000 titles, each belonging to 1 to 3 classes. The classes and corresponding number of titles are shown in Table 1:

TABLE 1. Headlines samples distribution.

Label	Number	Label	Number
International	6904	Pets	11203
Travel	29846	Military	12515
Technology	43922	Real estate	9934
Astrology	6153	Family	18428
Sports	37188	History	31413
Social	80502	Health	37373
Politics	39427	Game	7931
Food	46008	Finance	41177
Fashion	20455	Entertainment	51573
Education	24662	Design	7931
Culture	33982	Animation	11623
Collect	3589	Workplace	3749
Agriculture	14898	Childcare	20896
Automatic	27495	Other	15150

Baidu knows that as a prominent Chinese question-and-answer platform, it has massive Chinese question-and-answer data, so Baidu officially collected the question-and-answer data in 2019, established a Baidu question-and-answer data set, and released it(https://doi.org/10.5281/zenodo.3402023). The data set includes questions, question details, answers, and classes. Most of the question details and answers are long texts. In order to evaluate the multilabel classification ability of this article's model for Chinese short text, questions and classes were extracted to build a question classification data set. The question classification data set includes 570,000 questions in 24 classes, each belonging to 1 to 3 classes. The classes and corresponding number of questions are shown in Table 2.

TABLE 2. Questions samples distribution.

Label	Number	Label	Number
Education	171906	Fitness	33707
Entertainment	12471	Jewelry	10351
Network	152163	Accommodation	1411
Beauty	1025	Transportation	1325
Life	59863	Sports	20305
Food	18007	Study abroad	2853
Commercial	103468	Beverages	1915
System malfunction	21856	Event	1001
Electronic digital	19603	Automotive	2056
Communication	18186	Vocal	1365
Tourism	932	Hair care	1268
Culture	31960	Photography	1067

Before starting the testing, Chinese text needs to be preprocessed, which mainly includes text cleaning, word segmentation, and stop words removal. Among them, text cleaning refers to removing noise from the text. Word segmentation is the process of dividing continuous Chinese text into meaningful words, which also facilitates the extraction of pinyin information. Stop word removal involves deleting meaningless high-frequency words based on a predefined stop word list (such as Baidu's stop word list).

B. EVALUATION INDEX

We use average precision (P), average recall (R), and F1 value ($F1$) as evaluation indicators. The calculation formula is as shown in Equations (10)-(12):

$$P = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FP_i} \tag{10}$$

$$R = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i + FN_i} \tag{11}$$

$$F1 = \frac{2 \times P \times R}{P + R} \tag{12}$$

where is TP_i the number of samples FP_i that are predicted to be in i the class and are in the class, is i the number of samples FN_i that are predicted to be in i the class but do not belong to the class, and is the number of samples i that are predicted to not belong to i the class and actually do not belong to i the class.

C. EXPERIMENT SETTINGS

The model proposed in this paper is designed and implemented using Python, and the experiment is carried out on NVIDIA RTX3070 GPU; the operating system is ubuntu 16.04, and the memory is 16G. The experimental parameters are shown in Table 3:

TABLE 3. Parameter setting.

Parameter	Number
Epoch	20
Batch_size	64
Encoder layers	12
Attention headcount	12
Dropout	0.5
Learning_rate	1e-4

D. EXPERIMENT RESULTS and ANALYSIS

In order to evaluate the performance of the model proposed in this article, we compare it with the following models:

XML-CNN [12]: Based on the CNN model, it fully utilizes position information by introducing dynamic maximum pooling operations.

SGM [24]: The SGM model achieves multilabel text classification by considering label correlation and using the attention mechanism to obtain critical information about the text.

Seq2set [25]: Based on the SGM model, the introduction of a set decoder module fully leverages the unordered nature of labels and decreases the model's reliance on label sequences.

GUDN [26]: Utilizes a guidance network containing two guidance modules and two loss functions to guide BERT to extract features.

The results are shown in Table 4-5:

TABLE 4. Headlines classification results.

model	*P* (%)	*R* (%)	*F1* (%)
XML-CNN	84.35	81.09	82.69
SGM	86.05	83.21	84.61
SEQ2SET	86.99	84.02	85.48
GUDN	87.30	84.13	85.69
OUR	88.13	86.62	87.37

From Table 4 and Table 5, GUDN performs better than other comparison models, but compared to the method we proposed, there is still a gap of 1.68% and 1.47% in F1 score. Experimental results show that compared with the comparative model, the method proposed in this paper can extract semantic information more effectively and enhance the learning ability of text features, thereby achieving superior classification performance. Figure 4 shows the precision of the model for classifying classes with limited samples. From Figure 4, the proposed model has a better learning ability than the comparison model for class features with small samples. It is hypothesized that this result is tied to the GAN model's training approach, enabling it to effectively tackle issues caused by uneven sample distribution leading to poor classification.

E. ABLATION EXPERIMENT

An ablation experiment was set up to verify each part's importance in the model. The comparison model is as follows:

TABLE 5. Questions classification results.

model	*P* (%)	*R* (%)	*F1* (%)
XML-CNN	77.81	74.63	76.19
SGM	78.49	74.42	76.40
SEQ2SET	78.85	76.26	77.53
GUDN	79.27	78.91	79.09
OUR	81.41	79.72	80.56

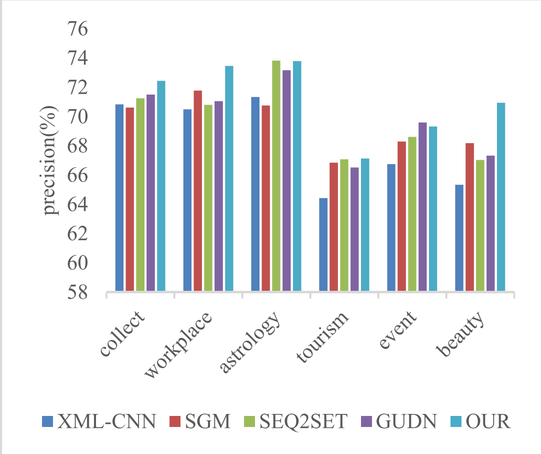


FIGURE 4. Classification results of small sample classes.

Without pinyin (WP): Only use the BERT model combined with the GAN model without adding pinyin embedding.

Without multiple generators (WMG): Connect a single generator to the last hidden layer of the BERT model.

Without improved sampling method (WISM): Use traditional sampling to extract negative samples.

The experimental results are shown in Table 6.

TABLE 6. Influence of each part.

model	Headlines dataset			Questions dataset		
	P (%)	*R* (%)	*F1* (%)	*P* (%)	*R* (%)	*F1* (%)
WP	86.62	86.24	86.43	79.72	78.55	79.13
WMG	87.29	84.63	85.94	79.22	77.97	78.59
WISM	87.05	85.42	86.23	80.27	78.43	79.34
OURS	88.13	86.62	87.37	81.41	79.72	80.56

Experimental results indicate that removing pinyin embeddings, multiple generators, and the improved sampling will diminish classification performance. Removing pinyin embeddings decreased the model F1 scores by 0.94% and 1.43%. The results confirm the importance of pinyin embeddings for Chinese semantic model; removing multiple generators reduced the model F1 scores by 1.43% and 1.97%. The decrease in F1 score shows that the distribution learning ability of a single generator for actual data is inferior to multiple generators, which hinders the learning of features by the GAN model; removing improved sampling results in a decrease in the model's F1 scores by 1.14% and 1.22%. Implementing improved sampling not only enhances the model's classification performance but also aids in reducing training steps.

V. CONCLUSION

We proposed a method for Chinese multilabel short text classification based on GAN, BERT and pinyin embedding. This method first enhances the learning of Chinese semantic features through pinyin embedding, thereby reducing the impact of sparse Chinese short-text features; second, by connecting multiple hidden layers of the BERT model to the generators, it incorporates diverse semantic information in the generated samples. Improve the probability of being similar to the actual sample distribution so that the discriminator can better learn the feature distribution; third, by improving the sampling method, we can improve the problem of convergence difficulties caused by positive and negative samples being too similar. It was verified on the data set, and the results were better than the comparative models, which verified the superiority of the model proposed in this article in the Chinese multilabel short text classification task.

In future work, when using multiple generators, we can consider treating the fake samples generated by each generator as independent samples, which may ensure that the negative samples are meaningful, thus further reducing the issue of the discriminator being unable to backpropagate effective gradients. Additionally, we could explore language models beyond BERT that offer enhanced support for Chinese, such as ERNIE, to facilitate a deeper understanding of semantic information.

REFERENCES

[1] T. Mikolov, A. Deoras, S. Kombrink, L. Burget, and J. Černocký, ''Empirical evaluation and combination of advanced language modeling techniques,'' in *Proc. Interspeech*, Aug. 2011, pp. 605–608.

[2] M.-L. Zhang and Z.-H. Zhou, ''ML-KNN: A lazy learning approach to multi-label learning,'' *Pattern Recognit.*, vol. 40, no. 7, pp. 2038–2048, Jul. 2007.

[3] Y. Zhou, J. Li, J. Chi, W. Tang, and Y. Zheng, ''Set-CNN: A text convolutional neural network based on semantic extension for short text classification,'' *Knowl.-Based Syst.*, vol. 257, Dec. 2022, Art. no. 109948.

[4] S. Bodapati, H. Bandarupally, R. N. Shaw, and A. Ghosh, ''Comparison and analysis of RNN-LSTMs and CNNs for social reviews classification,'' in *Advances in Applications of Data-Driven Computing*. Cham, Switzerland: Springer, 2021, pp. 49–59.

[5] T. Yang, L. Hu, C. Shi, H. Ji, X. Li, and L. Nie, ''HGAT: Heterogeneous graph attention networks for semi-supervised short text classification,'' *ACM Trans. Inf. Syst.*, vol. 39, no. 3, pp. 1–29, Jul. 2021.

[6] X. Chen, L. Xu, Z. Liu, M. Sun, and H. Luan, ''Joint learning of character and word embeddings,'' in *Proc. Innov. Appl. Artif. Intell. Conf.*, 2015, pp. 1236–1242.

[7] S. Cao, W. Lu, J. Zhou, and X. Li, ''cw2vec: Learning Chinese word embeddings with stroke n-gram information,'' in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 5053–5061.

[8] X. Tang, R. Zhu, T. Sun, and S. Wang, ''Moto: Enhancing embedding with multiple joint factors for Chinese text classification,'' in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 2882–2888.

[9] Z. Zhang, X. Han, Z. Liu, X. Jiang, M. Sun, and Q. Liu, ''ERNIE: Enhanced language representation with informative entities,'' in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 1441–1451.

[10] X. Sun, Z. Liu, and X. Huo, ''Six-granularity based Chinese short text classification,'' *IEEE Access*, vol. 11, pp. 35841–35852, 2023.

[11] Y. Ma, X. Liu, L. Zhao, Y. Liang, P. Zhang, and B. Jin, ''Hybrid embedding-based text representation for hierarchical multi-label text classification,'' *Expert Syst. Appl.*, vol. 187, Jan. 2022, Art. no. 115905.

[12] J. Liu, W.-C. Chang, Y. Wu, and Y. Yang, ''Deep learning for extreme multi-label text classification,'' in *Proc. 40th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Aug. 2017, pp. 115–124.

[13] Z. Chen, S. Li, L. Ye, and H. Zhang, ''Multi-label classification of legal text based on label embedding and capsule network,'' *Int. J. Speech Technol.*, vol. 53, no. 6, pp. 6873–6886, Mar. 2023.

[14] H. Fei, Y. Zhang, Y. Ren, and D. Ji, ''Latent emotion memory for multi-label emotion classification,'' in *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 5, pp. 7692–7699.

[15] S. Yan, ''Enhancing deep learning-based multi-label text classification with capsule network,'' *J. Phys., Conf. Ser.*, vol. 1621, no. 1, Aug. 2020, Art. no. 012037.

[16] J. Zhou, J. Wang, and G. Liu, ''Multiple character embeddings for Chinese word segmentation,'' in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics, Student Res. Workshop*, 2019, pp. 210–216.

[17] B. Liu, ''GCN-BERT and memory network based multi-label classification for event text of the Chinese government hotline,'' *IEEE Access*, vol. 10, pp. 109267–109276, 2022.

[18] X. Wang and L. Guo, ''Multi-label classification of Chinese rural poverty governance texts based on XLNet and bi-LSTM fused hierarchical attention mechanism,'' *Appl. Sci.*, vol. 13, no. 13, p. 7377, Jun. 2023.

[19] D. Croce, G. Castellucci, and R. Basili, ''GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples,'' in *Proc. 58th Annu. Meeting Assoc. Comput. Linguistics*, 2020, pp. 2114–2119.

[20] T. Auti, R. Sarkar, B. Stearns, A. K. Ojha, A. Paul, M. Comerford, J. Megaro, J. Mariano, V. Herard, and J. P. McCrae, ''Towards classification of legal pharmaceutical text using GAN-BERT,'' in *Proc. 1st Comput. Soc. Resp. Workshop 13th Lang. Resource Eval. Conf.*, 2022, pp. 52–57.

[21] S. Shehnepoor, R. Togneri, W. Liu, and M. Bennamoun, ''ScoreGAN: A fraud review detector based on regulated GAN with data augmentation,'' *IEEE Trans. Inf. Forensics Security*, vol. 17, pp. 280–291, 2022.

[22] M. Li, K. Yin, and M. Wang, ''Ptr4BERT: Automatic semisupervised Chinese government message text classification method based on transformer-based pointer generator network,'' *Adv. Multimedia*, vol. 2022, pp. 1–11, Aug. 2022.

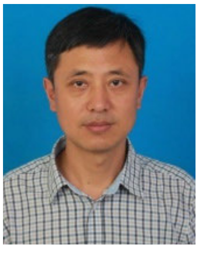
[23] R. Chiong, G. S. Budhi, S. Dhakal, and F. Chiong, ''A textual-based featuring approach for depression detection using machine learning classifiers and social media texts,'' *Comput. Biol. Med.*, vol. 135, Aug. 2021, Art. no. 104499.

[24] P. Yang, X. Sun, W. Li, S. Ma, W. Wu, and H. Wang, ''SGM: Sequence generation model for multi-label classification,'' in *Proc. COLING*, 2018, pp. 3915–3926.

[25] P. Yang, F. Luo, S. Ma, J. Lin, and X. Sun, ''A deep reinforced sequence-to-set model for multi-label classification,'' in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 5252–5258.

[26] Q. Wang, J. Zhu, H. Shu, K. O. Asamoah, J. Shi, and C. Zhou, ''GUDN: A novel guide network with label reinforcement strategy for extreme multi-label text classification,'' *J. King Saud Univ. Comput. Inf. Sci.*, vol. 35, no. 4, pp. 161–171, Apr. 2023.

JINPENG BAI received the B.S. degree from Hebei Normal University of Science and Technology, in 2018. He is currently pursuing the M.S. degree with the School of Cyber Security and Computer, Hebei University, Baoding, Hebei, China. His research interests include natural language processing and pattern recognition.



XINFU LI received the B.S. and M.S. degrees in electronic information engineering from Hebei University, Hebei, China, in 1993 and 1999, respectively, and the Ph.D. degree in intelligent electrical engineering from Hebei University of Technology, Tianjin, China, in 2004. He is currently a Professor with the School of Cyber Security and Computer, Hebei University. His main research interest includes image and text processing.