Check for updates

Latest updates: https://dl.acm.org/doi/10.1145/3522736

PDF Download
3522736.pdf
12 March 2026
Total Citations: 3
Total Downloads: 251

Published: 12 November 2022
Online AM: 15 March 2022
Accepted: 25 February 2022
Revised: 10 February 2022
Received: 06 August 2021

Citation in BibTeX format

RESEARCH-ARTICLE

AIP: A Named Entity Recognition Method Combining Glyphs and Sounds

BO LIU, Massey University, Palmerston North, MWT, New Zealand

ZHUO SU, Beijing University of Technology, Beijing, China

GUANGZHI QU, Oakland University, Rochester, MI, United States

Open Access Support provided by:

Beijing University of Technology

Massey University

Oakland University

AIP: A Named Entity Recognition Method Combining Glyphs and Sounds

BO LIU, Faculty of Information Technology, Beijing University of Technology, Beijing, China and School of Mathematical and Computational Sciences, Massey University, Palmerston North, New Zealand
ZHUO SU, Faculty of Information Technology, Beijing University of Technology, Beijing, China
GUANGZHI QU, Computer Science and Engineering Department, Oakland University, Rochester, MI, USA

In recent years, a large number of Chinese electronic texts have been produced in the process of information construction in various fields. Identifying specific entities in these electronic texts has become a major research focus. Most existing research methods use radicals to extract the glyph features of Chinese characters but have seen its limitation. This paper extracts the features of Chinese characters from three aspects: glyph features, phonetic features, and character features, and improves conventional feature extraction methods for each kind of feature. A new named entity recognition method (AIP) is proposed by transforming Chinese characters into corresponding images for glyph feature extraction, dividing pinyin into initials, vowels, and tones for phonetic feature extraction, and fine-tuning the A Lite Bert model for character feature extraction to improve the performance of the model. This paper compares the performance of the AIP model and mainstream neural network models on Chinese named entity recognition tasks on commonly used data sets and the data sets in specific domains. The results showed that AIP achieved better results than the related work. The F1 values on the two data sets are 94.4% and 80.5%, respectively, which validates the model's versatility.

CCS Concepts: • **Computing methodologies → Natural language processing**;

Additional Key Words and Phrases: Named entity recognition, glyph features, phonetic features

ACM Reference format:
Bo Liu, Zhuo Su, and Guangzhi Qu. 2022. AIP: A Named Entity Recognition Method Combining Glyphs and Sounds. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 21, 6, Article 127 (November 2022), 14 pages.
https://doi.org/10.1145/3522736

1 INTRODUCTION

In recent years, the rapid development of the Internet and the gradual acceleration of the construction of an information society has put forward higher requirements for the modernization system of various industries. As part of the construction of the information system of various industries,

This work is supported by National Natural Science Foundation of China (62076015).
Authors' addresses: B. Liu, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China and School of Mathematical and Computational Sciences, Massey University, Palmerston North 4472, New Zealand; email: liubo03@gmail.com; Z. Su, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; email: sz1995424@163.com; G. Qu (corresponding author), Computer Science and Engineering Department, Oakland University, Rochester, MI 48309, USA; email: gqu@oakland.edu.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2022 Association for Computing Machinery.
2375-4699/2022/11-ART127 $15.00
https://doi.org/10.1145/3522736

Fig. 1. The evolution of the Chinese word “患”.

the number of relevant Chinese electronic texts is increasing. How to extract some key information contained in Chinese electronic texts through technical means has important research value for the information and decision-making assistance of various industries. With the development of information technology, named entity recognition based on natural language processing has become a research hotspot in academia.

Chinese characters, as a kind of hieroglyph, have gone through a long development process. A very important point that differentiates Chinese from English characters is that Chinese contains rich glyph information which could be used for feature extraction.

Due to the particularity of Chinese characters that are different from English, many scholars have begun to study how to better capture the characteristics of Chinese characters. Some scholars have proposed to use the radicals of Chinese characters for Chinese named entity recognition tasks [1]. However, this method only captures a part of glyph features of Chinese characters, thus it is hard to capture the whole glyph features of Chinese characters thoroughly. At the same time, if the pinyin is coded as a whole, the pronunciation characteristics of Chinese characters are not fully reflected.

For example, in the Figure 1, the word “患” in Chinese, which can be seen from the appearance of the word “患” invented by the ancients in the early days, was imitated by the ancients observing the scene in reality. The word “患” is a pictophonetic character and related compound words. It means to penetrate from the heart, with the meaning of worry as piercing. If only the radical features are extracted, the radical of the word “患” is “心”, which only reflects the heart, but does not reflect the original meaning of the word “患”.

Each Chinese character has its pinyin, which is an important factor reflecting the pronunciation of Chinese characters. If the pinyin is coded according to the composition rules of pinyin, it can reflect the pronunciation characteristics of Chinese characters to a certain extent. For example, “啊” and “哦” are both modal auxiliary words. One of their pinyin is “á” and the other is “ó”. Their pinyin composition is all single vowels, and the tones are all two tones, expressing the meaning of exclamation.

80% of Chinese characters are phonetic and semantic compound words, and their phonetic and semantic parts provide a wide range of meanings [18]. How to better extract features from the components of Chinese characters and learn semantic representations has become the main research topic of this article.

This paper proposes a new model AIP to better capture the characteristic information of words and improve the recognition efficiency. The glyph and phonetic information of the characters are extracted in a brand-new way, and the four layers that are the most effective for named entity recognition by ALBERT are spliced and optimized, and finally, a brand-new network structure is formed to recognize Chinese text entities.

The contributions of this article are summarized as follows:

(1) Chinese characters are a long-standing language and a kind of pictographic character. Traditional feature extraction methods such as radical strokes cannot fully capture the glyph features of Chinese characters. This article converts Chinese characters into images

and adopts the denser DenseNet model, which can fully capture the glyph features of Chinese characters from a more vivid angle. Since the pre-trained model has undergone training on a large-scale dataset, which has better generality, passing the formed images through the pre-training model solves the problem of lacking Chinese training data to a certain extent.

(2) If the entire pinyin of Chinese characters is coded in the phonetic feature extraction of Chinese characters, the similarity between pinyin pronunciations cannot be thoroughly utilized. The similarity between pronunciations can also capture the correlation between characters to a certain extent. This article proposes a new encoding method where the pinyin is not coded as a whole block, but unpacked into initials, vowels, and tones according to the pronunciation components for coding, which better captures the phonetic features of Chinese characters.

(3) Different BERT layers can obtain different levels of semantic and syntactic information. For the named entity recognition task performed by BERT, this work chooses the lighter ALBERT. The cropping and fine-tuning of the ALBERT model not only has little effect on the semantic ability of the model, but also has some improvements on it, and at the same time, the number of model parameters and reasoning time are greatly reduced. This paper selects and splices the four important layers in the named entity recognition task in ALBERT, realizes the fine-tuning of the ALBERT model, and improves the performance of the model.

(4) The good recognition results of the model on the Chinese named entity recognition data set were verified on multiple data sets, and the effects of commonly used models were compared on different training sets and test sets.

In summary, the advantage of the proposed model lies in the analysis of the characteristics of Chinese characters from three aspects, i.e., glyph features, phonetic features, and character features, and the potential semantic information of Chinese characters is well extracted.

2 RELATED WORK

In the field of named entity recognition, early named entity methods were based on linguists manually setting rule templates to perform pattern matching, thereby classifying entity types [2]. Later, as related models in the field of machine learning were proposed, various methods of machine learning began to be applied in this field. The named recognition problem is generally regarded as a sequence labeling problem [26] with machine learning methods, where the conditional probability of the label sequence under the condition of the input sequence is usually modeled, and then the collected annotated corpus is used for feature learning and model parameter training. Finally, in the prediction stage, the trained model is used to identify the named entities in the unlabeled corpus. Traditional machine learning methods mainly use the Hidden Markov model and conditional random fields. Usually, a node is used to represent one or a group of random variables, and the edges between nodes represent the probability relationship between variables. Fu et al. performed Chinese named entity recognition based on lexical HMM, unifying unknown words and known words into the task of marking the sequence of known words [3]. CRF could define a wider feature set and set any weight value, which is improved compared to HMM. Many scholars have conducted research on the task of Chinese named entity recognition based on the CRF model [4, 5]. With the further development of deep learning, some scholars began to apply the RNN structure to the field of named entity recognition and proposed a special RNN structure: LSTM structure [6]. Zhiheng Huang et al. proposed a two-way LSTM-CRF model for sequence labeling. For the first time, the BiLSTM+CRF model was used for sequence labeling tasks and the best results were obtained, which opened the prelude to the era of deep learning for named entity recognition [7]. Based on the idea of BiLSTM-CRF, many scholars have proposed various improved versions based on the structure of BiLSTM-CRF. For example, Xuezhe Ma and Eduard Hovy proposed an end-to-end two-way LSTM-CNNS-CRF model at the ACL 2016 Conference [8].

In ACL 2018, Yue Zhang and Jie Yang proposed a Chinese named entity recognition method based on lattice LSTM, introduced vocabulary enhancement to solve the Chinese named entity recognition problem, and used dictionary information to enhance the effect of named entity recognition [9]. Some scholars use CNN structure instead of RNN to do Chinese named entity recognition to solve the problem that RNN structure cannot be processed in parallel and cannot fully utilize GPU. For example, Tao Gui, Ruotian Ma, et al. proposed a CNN-based LR-CNN model with a rethinking mechanism [10], and Yuying Zhu et al. proposed a CAN network based on convolutional attention networks and gated recurrent units for Chinese named entity recognition [11]. Liu, Li, et al. compared named entity recognition methods based on words and phrases. The experimental results showed that the word-based named entity recognition method has better performance than the phrase-based ones [12, 13]. Lu, Dong, and others have adopted a word-based entity recognition scheme [1, 14]. In recent years, the mainstream method of named entity recognition is to use the pre-training model represented by BERT [15]. Yu et al. used BERT-BiLSTM-CRF to obtain better experimental results, surpassing the previous mainstream methods [16].

In the aspect of Chinese characters, some scholars have also conducted certain research on radicals. Sun et al. proposed to use a feedforward neural network and used Chinese radical features as a supervised tag training word embedding [17]. Y. Li et al. presented a model that used character glyphs to enhance character feature representation [18]. Y. Wu et al. depicted a radical neural network RCBC to use the radical features of Chinese characters for Chinese named entity recognition [19]. X. Shi et al. represented a "radical embedding" technique and adopted Word2Vec to pre-train the radical vector [20]. In addition, some researchers extract the glyph features of Chinese characters from their graphic aspects. Meng et al. proposed GLYCE, the glyph-vectors for Chinese character representations. They treated Chinese characters as images and used CNNs to obtain their representations [27]. FGN was proposed by Xuan et al. On the one hand, a new CNN structure was proposed, called CGS-CNN. On the other hand, a method with a sliding window and Slice-Attention to fuse the BERT representation and glyph representation for a character was provided [28]. Recently, some scholars paid attention to not only the glyph features but also the pinyin features of Chinese characters. Sun et al. proposed ChineseBERT, starting from the two characteristics of Chinese characters themselves, and integrating the glyph and pinyin information of Chinese characters into the pre-training process of Chinese corpus [29].

3 METHOD

3.1 AIP Model

In order to address the existing problems, this paper proposed a novel method named **AIP (ALBERT-IMAGE-PHONETIC** MODEL) for Chinese named entity recognition. This section first elaborates the overview of the AIP model (as seen in Figure 2), and then separately illustrates the three parts of the model, including the extraction of character features, glyph features, and phonetic features.

The whole model is composed of four parts, including three parallel branches (the glyph feature extraction part, the phonetic feature extraction part, the character feature extraction part), and the final model prediction part.

First, the texts of Chinese characters are converted into images. After conversion, we reconstruct the image size so that each Chinese character corresponds to an image with the size 224*224. Then,

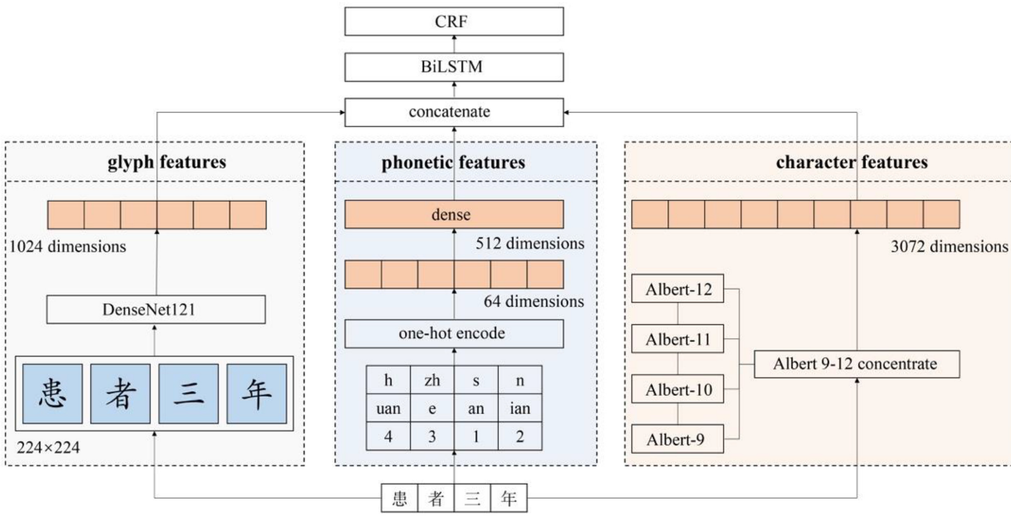

Fig. 2. Overview of the AIP model.

the processed Chinese character images were passed through the pre-trained DenseNet121 model and each Chinese character image was converted into a 1024-dimensional vector.

Second, each character's pinyin is divided into initials, finals, and tones according to its components and one-hot encoding is performed according to their respective types. Each Chinese character is converted into a 64-dimensional vector after one-hot encoding. To prevent the feature from being ineffective due to too small dimensions, we have added a dense layer, so that each Chinese character will be converted into a 512-dimensional vector after extracting the phonetic feature.

Third, the text information is converted into 3072-dimensional vectors one by one by fine-tuning the pre-trained model, that is, fusing the last four layers of the ALBERT model.

Finally, the feature vectors obtained from the three parts are spliced, then the BiLSTM-CRF model is adopted to train the processed vectors, and finally the entity labeling is obtained.

3.2 DenseNet

Many researchers have done in-depth development on the semantic representation of Chinese, mainly focusing on the radicals of Chinese characters. But we found that the similarity of Chinese characters themselves is more correlated than the radicals, so this study intends to capture the correlation between similar characters by taking each Chinese character as a whole image. However, the traditional CNN model is not ideal for extracting local features [21], so this study intends to use the closely connected convolutional network DenseNet to capture the features in the Chinese character feature map while suppressing the problem of overfitting to a certain extent.

DenseNet [22] proposes a denser connection mechanism: that is, all layers are connected to each other. Specifically, each layer accepts all the previous layers as its additional input. If the convolutional network has a shorter connection between the layer close to the input and the layer close to the output, then training through this convolutional network can be more in-depth and effective. Figure 3 shows the dense connection mechanism of DenseNet. Each layer is connected with all previous layers in the channel dimension and used as the input of the next layer. For an l layers network, DenseNet contains a total of $l(l+1)/2$ connections. As shown in Figure 3, the

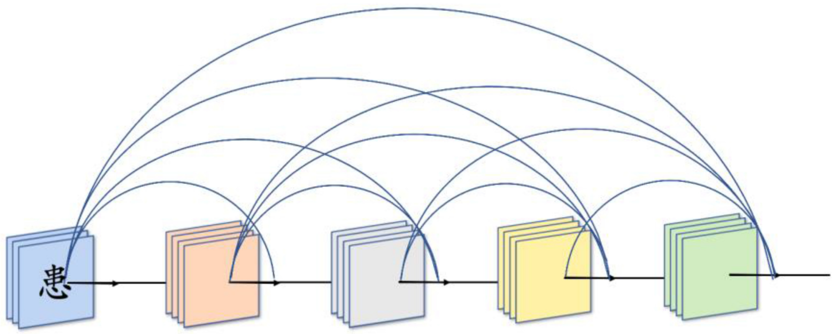

Fig. 3. DenseNet network structure [22].

outstanding advantage of DenseNet is to alleviate the disappearance of gradients, strengthen the propagation of features, and greatly reduce the number of parameters [22].

The corresponding formula (1) is used to express:

$$x_L = H_l\left([x_0, x_1, \ldots, x_{l-1}]\right) \tag{1}$$

The l layer accepts the feature maps of all previous layers $x_0, x_1, \ldots, x_{l-1}$ as input. Among them $[x_0,\ x_1, \ldots, x_{l-1}]$ refers to the splicing of feature maps from layer 0 to layer $l-1$. H_l is defined as a compound operation of three consecutive operations, including a series of **batch normalization (BN), modified linear unit (RELU)**, and 3*3 convolution operation. In this research, the characters in the Chinese character library are drawn one by one to form a 224*224 pixel picture, and then feature extraction is performed through the tightly connected DenseNet model to form a 1024-dimensional vector for subsequent stitching and input.

3.3 Albert

BERT applications have achieved satisfactory results in various natural language processing projects and are widely used in Chinese named entity recognition tasks. However, the pre-training model BERT has the problem of excessive parameters. In order to reduce the parameters required in BERT, scholars proposed the ALBERT model, which requires fewer parameters and is more suitable for engineering tasks compared with BERT [23].

The results obtained at different layers of BERT often represent different grammatical and semantic information. Studies have shown that fine-tuning BERT on the text task does not greatly affect the results of the experiment and can make better use of BERT for text tasks. Choosing different layers of models has different effects in text tasks, and the top layer has the best effect [24].

Similar to BERT, the results obtained by different layers of the lighter ALBERT also output different semantic information. This paper selects the model ALBERT that requires fewer parameters and selects the last four layers of ALBERT as the result of fine-tuning, to extract character features for this recognition task.

The maximum sentence length of ALBERT is 512. We segmented the experimental text according to the maximum length of 500. If the sentence length is greater than 500, we segmented the sentence with a period, question mark, exclamation mark, etc. to form multiple training sentences, and added "[SEP]" mark in front of the sentence and "[CLS]" mark at the end of the sentence according to the requirements of ALBERT input sentence, and finally input it into the fine-tuned ALBERT model for character feature extraction. Since we selected the last four layers in ALBERT, the training data will form a 768*4 = 3072 dimensional vector after passing ALBERT.

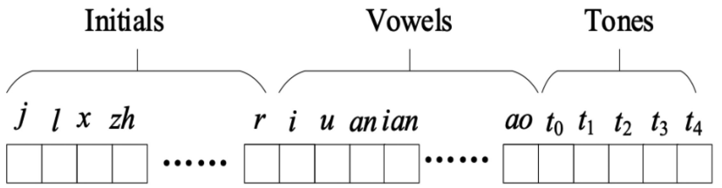

Fig. 4. Pinyin coding diagram.

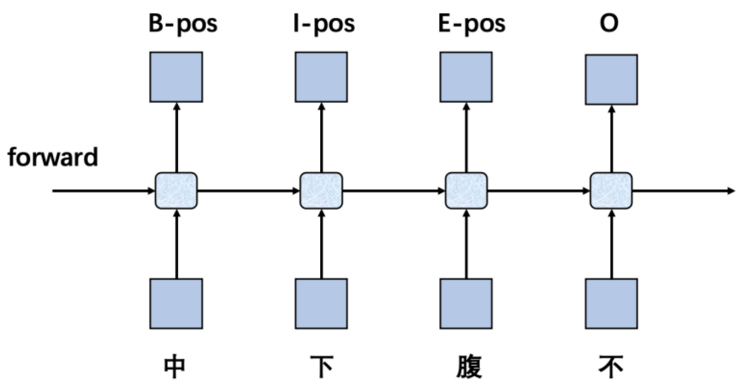

Fig. 5. LSTM model [7].

3.4 Pinyin

Pinyin represents the pronunciation of Chinese characters. Pinyin plays an irreplaceable role in the composition of Chinese characters. The process of literacy is a process of learning pinyin. Through pinyin, we can recognize new characters and gain the ability to learn Chinese characters independently. Pinyin can be divided into initials, vowels, and tones. The three parts are combined to form the pronunciation of a Chinese character. Taking the Chinese character "患" as an example, its pinyin is "huàn", the initial is "h", the vowel is "uan", and the tone is "`". Many Chinese named entity recognition tasks used pinyin as a feature. If the pinyin is coded as a whole, the pronunciation characteristics of the Chinese characters themselves will be ignored. Therefore, this research starts from the relevance of the pronunciation of Chinese characters, divides the pinyin coding into initials, vowels, and tones for coding, so as to better restore the pronunciation of Chinese characters and better capture the correlation between the pronunciation of Chinese characters, and finally forms a 64-dimensional vector after encoding, as shown in Figure 4.

3.5 BiLSTM-CRF

The RNN model was applied to named entity recognition tasks in 1997. Based on its characteristics, RNN can predict the current output based on long-distance elements [7]. But in practice, it tends to favor the latest input of the sequence [25]. Later scholars proposed the LSTM network, as shown in Figure 5. The introduction of the gate memory mechanism, using memory cells, has the powerful function of capturing long-distance dependence.

The specific formula is as follows:

$$i_t = \sigma\left(W_i x_t + U_i h_{t-1} + b_i\right) \tag{2}$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \tag{3}$$

$$\widetilde{C}_t = tanh\left(W_c x_t + U_c h_{t-1}\right) \tag{4}$$

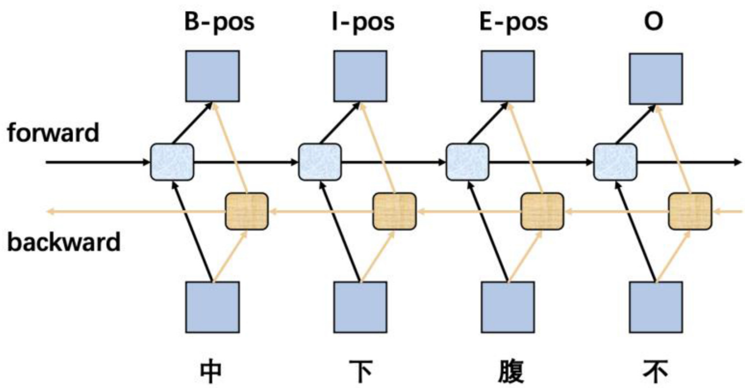

Fig. 6. BiLSTM model [7].

$$o_t = \sigma \left(W_o x_t + U_o h_{t-1} + b_o \right) \tag{5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \tag{6}$$

$$h_t = o_t \odot tanh \left(C_t \right) \tag{7}$$

Here σ is the logistic regression function, and i, f, and o are input gate, forget gate, and output gate, respectively. x is the current input, W and U are the weight matrix, C is the cell state, and h is the hidden state.

The LSTM model can better capture the long-distance dependency but cannot capture the information from the back to the front, so BiLSTM is often used to capture the dependency in the front and back directions. BiLSTM [7] is composed of forward LSTM and backward LSTM. To make better use of past and future information, we will splice the left-to-right LSTM network and the right-to-left LSTM network to get a two-way LSTM network, as shown in Figure 6.

The CRF [30] model could learn the contextual dependencies of sentences, and the introduction of the CRF model after the BiLSTM model can strengthen the constraints to ensure that the final recognition and prediction effect is more effective. For example, the beginning of named entity recognition should be b-entity instead of i-entity, and the middle should be i-entity instead of b-entity. Adding a layer of CRF model after the BiLSTM model could effectively avoid these errors and add effective constraints to the final predicted label. The BiLSTM-CRF model is shown in Figure 7.

The CRF loss function is expressed by the following formula:

$$LossFunction = \frac{P_{realpath}}{P_1 + P_2 + \cdots + P_n} \tag{8}$$

The CRF loss function consists of two parts, the score of the real path and the total score of all paths. The score of the real path is the highest of all the paths. During the training process, the parameter values of the model will be continuously updated with the iteration of the training process, making the ratio of the real path larger and larger.

4 EXPERIMENTS

4.1 Data Sets

This article uses two data sets to train and test the proposed model. One is a part of the widely used MSRA data set, and the other is a field-specific data set: the medical named entity recognition

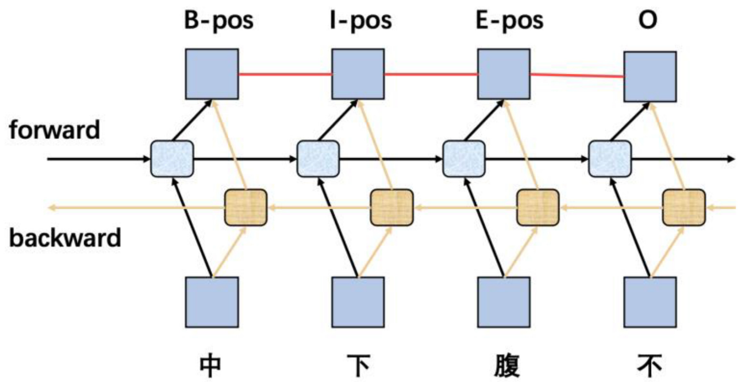

Fig. 7. BiLSTM-CRF model [7].

Table 1. The Distribution of Each Entity in the MSRA Data Set

Entity category	Number	Percentage
NS	6037	0.44
NT	4077	0.30
NR	3541	0.26

data set in the CCKS2020 Chinese electronic medical record medical entity and event extraction task.

The MSRA data set includes three naming types, NS (LOCATION), NR (PERSON), and NT (ORGANIZATION). This data set is labeled with BMES, in which *B* represents the beginning of an entity, *M* refers to the interior of an entity, *E* is the result of an entity, *O* represents an external element, and *S* denotes the word that can be used as an entity alone.

The CCKS2020 medical named entity recognition data set includes six named types: DIS(DISEASE), OPE(OPERATION), POS(POSITION), MED(MEDICINE), LAB(LABORATORY), and SCR(SCREEN). This data set is labeled with BIOES, in which *B* represents the beginning of an entity, *I* describes the interior of an entity, *E* is the result of an entity, *O* refers to an external element, and *S* represents the word that can be used as an entity alone.

The distribution of the training set entities of the two data sets is shown in Tables 1 and 2.

4.2 Evaluation Index

This experiment intends to use P (precision rate), R (recall rate), and F1 value to evaluate the results of this model and other comparable models on the data set. The specific calculation formulas are as follows:

$$P = \frac{TP}{TP + FP} \qquad (Precision)$$

$$R = \frac{TP}{TP + FN} \qquad (Recall)$$

$$F = \frac{2P \times R}{P + R} \qquad (F1)$$

Table 2. The Distribution of Each Entity in the CCKS2020 Data Set

Entity category	Number	Percentage
DIS	6004	0.30
POS	6782	0.34
MED	2731	0.14
LAB	1797	0.09
SCR	1466	0.07
OPE	1256	0.06

Table 3. Comparison of F1-Score on MSRA Data Set

Model	F1-score
BERT-BiLSTM-CRF	93.1
ALBERT-BiLSTM-CRF	92.7
Lattice-LSTM	90.1
ChineseBERT	93.6
AIP-BiLSTM-CRF	94.4

Where TP represents the number of correct predictions for positive cases, FP shows the number of prediction errors for negative cases, TN reveals the number of correct predictions for negative cases, and FN indicates the number of prediction errors for positive cases.

4.3 Experimental Result

In the experiment, the ratio of training samples to test samples is 10:1, and the training-related settings are as follows: batch_size is set to 32, the learning rate is set to le-4, the training optimizer is set to Adam, and the loss function is selected as the loss function of CRF.

To validate the performance of the proposed model, the ALBERT-BiLSTM-CRF model, the BERT-BiLSTM-CRF model, the ChineseBERT model, and the Lattice-LSTM model are used as the baselines. The ALBERT-BiLSTM-CRF model and the BERT-BiLSTM-CRF model convert text to vectors using pre-trained ALBERT and BERT models, respectively, and then use BiLSTM-CRF model to train the processed vectors and finally complete the entity tagging. The Lattice-LSTM [9] model uses a dynamic network structure fused with word and vocabulary information to solve the Chinese named entity recognition problem. The ChineseBERT [29] integrates the glyph and pinyin information of Chinese characters into the Chinese pre-training model to do Chinese named entity recognition tasks.

The named entity recognition method AIP-BiLSTM-CRF proposed in this paper is compared with the four baseline methods of the test dataset as shown in Table 3.

It can be seen from the above table that the proposed model is better than the commonly used NER models, including the BERT-BiLSTM-CRF model and the ALBERT-BiLSTM-CRF model on the MSRA data set. Meanwhile, the effect of our model is greatly improved compared to the Lattice-LSTM model. The F1 value of our model is 94.4, while the F1 values of the BERT-BiLSTM-CRF model, the ALBERT-BiLSTM-CRF model, the Lattice-LSTM model, and the ChineseBERT model are 93.1, 92.7, 90.1, and 93.6, respectively. The proposed model is 1.3% higher than the BERT-BiLSTM-CRF model, 1.7% higher than the ALBERT-BiLSTM-CRF model, 4.3% higher than the Lattice-LSTM model, and 0.8% higher than the ChineseBERT Model concerning F1 value. According to the

Table 4. The Performance of the AIP Model on the Test Set

Entity	Precision	Recall	F1-score	Support
B-NR	0.98	0.89	0.93	251
B-NS	0.92	0.87	0.89	314
B-NT	0.78	0.73	0.75	179
E-NR	0.97	0.86	0.92	251
E-NS	0.91	0.84	0.87	314
E-NT	0.88	0.77	0.82	179
M-NR	0.98	0.93	0.95	185
M-NS	0.89	0.69	0.78	165
M-NT	0.85	0.83	0.84	527
O	0.99	1.00	0.99	31466
S-NR	0.33	0.20	0.25	10
S-NS	1.00	0.31	0.47	39

Table 5. The Ablation Experiment Diagram of the AIP Model

Model	F1-score
ALBERT-BiLSTM-CRF	92.7
ALBERT(the last four layers)-BiLSTM-CRF	93.7
ALBERT(the last four layers)-glyph feature-BiLSTM-CRF	94.2
AIP-BiLSTM-CRF	94.4

experimental results, our model has achieved good performance through the comprehensive improvement of character glyph feature extraction, character phonetic feature extraction, and character feature extraction.

We further studied the recognition effect of the model proposed in this paper on different types of entities, and the effect is shown in Table 4.

As can be seen from Table 4, in the test set, except for the O entity, the recognition effect is the best on the NR entity, especially the M-NR entity, which reaches 0.95. In other entities, due to the small number of S-NR entities and S-NS entities trained, the recognition effect is not ideal, and the recognition effect of other entities is very good.

At the same time, to verify the character features extracted by ALBERT in the four layers of the fusion and the improvement effect after adding glyph features and character sound features, this experiment uses the ALBERT-BiLSTM-CRF model, the ALBERT (the last four layers) -BiLSTM-CRF model, the ALBERT (the last four layers) - glyph feature-BiLSTM-CRF model, and the AIP-BiLSTM-CRF model to perform ablation experiments, respectively. The ALBERT (the last four layers) -BiLSTM-CRF model improves the character feature extraction part based on the ALBERT-BiLSTM-CRF model, that is, the ALBERT model with four layers after fusion is used for character feature extraction. The ALBERT (the last four layers) -glyph feature-BiLSTM-CRF model adds glyph features on the basis of the ALBERT (the last four layers) -BiLSTM-CRF model. The AIP-BiLSTM-CRF model adds phonetic features to the ALBERT (the last four layers) -glyph feature-BiLSTM-CRF model. The experimental results are shown in Table 5.

It can be seen from the experimental results that compared with the conventional ALBERT-BiLSTM-CRF model, the ALBERT-BiLSTM-CRF model after the four layers after the integration of ALBERT's outstanding performance for text tasks has increased by 1.0%. After adding the glyph

Table 6. Comparison of F1-Score on the CCKS2020 Data Set

Model	F1-score
BERT-BiLSTM-CRF	77.8
ALBERT-BiLSTM-CRF	75.4
Lattice-LSTM	76.3
ChineseBERT	80.4
AIP-BiLSTM-CRF	80.5

feature on this basis, it increased by 0.5%, and after adding the phonetic feature, it increased by 0.2%. It can be seen that, based on the conventional model, the improvement of the character feature extraction process, the addition of the character shape feature, and the addition of the character sound feature have a relatively obvious improvement effect. Meanwhile, it is concluded from the experiment that the improvement of the character shape feature is greater than the that of the character sound feature.

To verify the versatility of the model, we conducted further experiments on the CCKS2020 data set in a specific field. The proposed model is compared with the four methods (BERT-BiLSTM-CRF, ALBERT-BiLSTM-CRF, Lattice-LSTM, and ChineseBERT) on the test dataset as shown in Table 6.

It can be seen from the experimental results that the AIP-BiLSTM-CRF model proposed in this paper also has a good effect on the dataset CCKS2020. The F1 value of this model reached 80.5, and the F1 values of the BERT-BiLSTM-CRF, ALBERT-BiLSTM-CRF, Lattice-LSTM, and ChineseBERT model are 77.8, 75.4, 76.3, and 80.4, respectively. In contrast, the AIP model has a significant improvement in the electronic medical record data set compared with the mainstream models. Meanwhile, we found that the performance of the AIP model compared with other models is more prominent on the professional Chinese electronic case data set with many entity categories. This further reflects that the improvement of character shape feature extraction, character sound feature extraction, and character feature extraction have played an important role in promoting the model effect. We further verified the performance of the proposed model on different types of entities in the CCKS2020 test set. The results are shown in Table 7.

It can be seen from the above table that the AIP model has an outstanding recognition effect on the CCKS2020 electronic medical record test set. In addition to the O entity, the recognition effect is the best on the OPE entity, especially the I-OPE entity that has reached an F1 value of 0.95. Except for the S-DIS entity, the F1 values of other entities are basically above 0.80. It can be seen that the model proposed in this paper also has a good effect on the data set in a specific field with good versatility.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have proposed the AIP method for Chinese named entity recognition. Starting from the glyph, phonetics, and character features of Chinese characters, we have made optimizations from three aspects. DenseNet is used to extract the glyph features, the initials, vowels, and tones are used to extract the phonetic features and the last four layers of ALBERT fusion is used to extract character features. Finally, the multi-directional feature extraction of Chinese characters is combined with BiLSTM-CRF for training. In the field of Chinese named entity recognition, it has achieved better results than mainstream models on two different data sets, which proves the effectiveness of this model. In addition, the improvement of each step was validated in the experiments. It is proved that the glyph features, phonetic features, and character features all play an important role in the task of Chinese named entity recognition. By better extracting these three features of

Table 7. Performance of AIP Model on the CCKS2020 Test Set

Entity	Precision	Recall	F1-score	Support
B-DIS	0.88	0.87	0.88	379
B-LAB	0.81	0.86	0.83	141
B-MED	0.91	0.96	0.93	176
B-OPE	0.91	0.90	0.91	94
B-POS	0.83	0.84	0.83	413
B-SCR	0.80	0.88	0.84	68
E-DIS	0.87	0.88	0.87	370
E-LAB	0.71	0.82	0.76	131
E-MED	0.88	0.96	0.92	171
E-OPE	0.91	0.96	0.93	89
E-POS	0.79	0.78	0.78	422
E-SCR	0.77	0.94	0.85	62
I-DIS	0.87	0.90	0.88	2107
I-LAB	0.82	0.90	0.86	293
I-MED	0.93	0.93	0.93	279
I-OPE	0.95	0.94	0.95	1164
I-POS	0.85	0.81	0.83	833
I-SCR	0.88	0.92	0.90	130
O	0.99	0.98	0.98	33951
S-DIS	0.00	0.00	0.00	0
S-LAB	0.75	1.00	0.86	3
S-POS	0.89	0.92	0.91	338

Chinese characters, good results can be obtained in the task of Chinese named entity recognition. Moreover, this research does not focus on a specific field and is versatile. In the future, we will experiment with the influence of different Chinese fonts on the model effect, and judge the relationship between the pictographic degree of the font and the model effect. In addition, changing the decoder part of the model may also have a significant improvement on the model.

REFERENCES

[1] C. Dong, J. Zhang, C. Zong, M. Hattori, and H. Di. 2016. Character-based LSTM-CRF with radical-level features for Chinese named entity recognition. In *Natural Language Understanding and Intelligent Applications*. Springer, Cham, 239–250.

[2] A. Mansouri, L. S. Affendey, and A. Mamat. 2008. Named entity recognition approaches. *International Journal of Computer Science and Network Security* 8, 2 (2008), 339–344.

[3] G. Fu and K. K. Luke. 2005. Chinese named entity recognition using lexicalized HMMs. *ACM SIGKDD Explorations Newsletter* 7, 1 (2005), 19–25.

[4] W. Chen, Y. Zhang, and H. Isahara. 2006. Chinese named entity recognition with conditional random fields. In *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*. 118–121.

[5] K. Liu, Q. Hu, J. Liu, and C. Xing. 2017. Named entity recognition in Chinese electronic medical records based on CRF. In *2017 14th Web Information Systems and Applications Conference (WISA)*. IEEE, 105–110.

[6] S. Hochreiter and J. Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.

[7] Z. Huang, W. Xu, and K. Yu. 2015. Bidirectional LSTM-CRF models for sequence tagging. *arXiv preprint arXiv*:1508.01991.

[8] X. Ma and E. Hovy. 2016. End-to-end sequence labeling via Bi-directional LSTM-CNNs-CRF. *arXiv preprint arXiv*:1603.01354.

[9] Y. Zhang and J. Yang. 2018. Chinese NER using Lattice LSTM. *arXiv preprint arXiv*:1805.02023.

[10] T. Gui, R. Ma, Q. Zhang, L. Zhao, Y. G. Jiang, and X. Huang. 2019. CNN-based Chinese NER with lexicon rethinking. In *IJCAI*. 4982–4988.

[11] Y. Zhu, G. Wang, and B. F. Karlsson. 2019. CAN-NER: Convolutional attention network for Chinese named entity recognition. *arXiv preprint arXiv*:1904.02141.

[12] Z. Liu, C. Zhu, and T. Zhao. 2010. Chinese named entity recognition with a sequence labeling approach: Based on characters, or based on words?. In *International Conference on Intelligent Computing*. Springer, Berlin, 634–640.

[13] H. Li, M. Hagiwara, Q. Li, and H. Ji. 2014. Comparison of the impact of word segmentation on name tagging for Chinese and Japanese. In *LREC*. 2532–2536.

[14] Y. Lu, Y. Zhang, and D. Ji. 2016. Multi-prototype Chinese character embedding. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. 855–859.

[15] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. 2018. BeERTPre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv*:1810.04805.

[16] X. Yu, W. Hu, S. Lu, X. Sun, and Z. Yuan. 2019. BioBERT based named entity recognition in electronic medical record. In *2019 10th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, 49–52.

[17] Y. Sun, L. Lin, N. Yang, Z. Ji, and X. Wang. 2014. Radical-enhanced Chinese character embedding. In *International Conference on Neural Information Processing*. Springer, Cham, 279–286.

[18] Y. Li, W. Li, F. Sun, and S. Li. 2015. Component-enhanced Chinese character embeddings. *arXiv preprint arXiv*:1508.06669.

[19] Y. Wu, X. Wei, Y. Qin, and Y. Chen. 2019. A radical-based method for Chinese named entity recognition. In *Proceedings of the 2nd International Conference on Big Data Technologies*. 125–130.

[20] X. Shi, J. Zhai, X. Yang, Z. Xie, and C. Liu. 2015. Radical embedding: Delving deeper to Chinese radicals. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 594–598.

[21] M. Zeng and N. Xiao. 2019. Effective combination of DenseNet and BiLSTM for keyword spotting. *IEEE Access*, 7, 10767–10775.

[22] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 4700–4708.

[23] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. 2019. Albert: A Lite BERT for self-supervised learning of language representations. *arXiv preprint arXiv*:1909.11942.

[24] C. Sun, X. Qiu, Y. Xu, and X. Huang. 2019. How to fine-tune BERT for text classification?. In *China National Conference on Chinese Computational Linguistics*. Springer, Cham, 194–206.

[25] Y. Bengio, P. Simard, and P. Frasconi. 1994. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks* 5, 2 (1994), 157–166.

[26] J. Li, A. Sun, J. Han, and C. Li. 2020. A survey on deep learning for named entity recognition. *IEEE Transactions on Knowledge and Data Engineering*.

[27] Y. Meng, W. Wu, F. Wang, X. Li, P. Nie, F. Yin, and J. Li. 2019. Glyce: Glyph-vectors for Chinese character representations. *arXiv preprint arXiv:1901.10125*.

[28] Z. Xuan, R. Bao, and S. Jiang. 2020. FGN: Fusion glyph network for Chinese named entity recognition. In *China Conference on Knowledge Graph and Semantic Computing*. Springer, Singapore, 28–40.

[29] Z. Sun, X. Li, X. Sun, Y. Meng, X. Ao, Q. He, and J. Li. 2021. ChineseBERT: Chinese pretraining enhanced by glyph and pinyin information. *arXiv preprint arXiv:2106.16038*.

[30] J. Lafferty, A. McCallum, and F. C. Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Received 6 August 2021; revised 10 February 2022; accepted 25 February 2022