Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **8**, Number **2**, May **2023**

Named Entity Recognition of Chinese Text Based on Attention Mechanism

Tong-Ping Shen*

School of Information Engineering
Anhui University of Chinese Medicine
Hefei, 230012, China
Graduate school
Angeles University Foundation
Angeles City, 2009, Philippines
shentp2010@ahtcm.edu.cn

Menchita Dumlao

Graduate school
Angeles University Foundation
Angeles City, 2009, Philippines
dumlao.menchita@auf.edu.ph

Qing-Quan Meng

School of Information Engineering
Anhui University of Chinese Medicine
Hefei, 230012, China
mqq1203@qq.com

Zhong-Hua Zhan

School of Information Engineering
Anhui University of Chinese Medicine
Hefei, 230012, China
58681333@qq.com

*Corresponding author: Tong-Ping Shen

Received December 20, 2022, revised March 9, 2023, accepted April 17, 2023.

Abstract. *The classical BiLSTM-CRF (Bidirectional Long Short-Term Memory-Conditional Random Field) model suffers from problems of polysemy, distraction of modeling at the coding layer, and lack of local spatial feature capture, which directly affect the recognition accuracy of text entities. In order to solve the above problems, this paper proposes an entity recognition structure combining BERT and MHA (Multi-Head Attention Mechanism).The BERT pre-training model adopts the full-word MASK model, which can effectively obtain contextual information and is more suitable for processing Chinese text. The processed text vectors from BERT layer are passed to BiLSTM layer to further obtain bidirectional text information, and then the semantic vectors are input to MHA layer to obtain information on temporal features, spatial features and character importance assignment weights of the input sequence, and finally the global optimal annotation sequence is obtained through CRF layer. The F_1 values of the integrated model reached 96.37% and 96.14% on the People's Daily and MSRA (Microsoft Research Asia) datasets, respectively, which verified the effectiveness of the method. The Chinese entity recognition model proposed in this paper can further improve the entity recognition effect and accuracy rate, and enhance the technical support for the subsequent work of natural language processing.*
Keywords: chinese text; entity recognition; multi-headed attention; BERT; datasets

1. Introduction.

Chinese text entity recognition focuses on identifying specific entities from unstructured and semi-structured texts, such as name, ethnicity, education, nationality, organization, etc. The Chinese text entity recognition research started at the MUC-6 conference in 1995 [1]. Sun et al. [2] proposed a statistical rule-based approach to automatically identify people's names in Chinese texts. English text can be divided by upper and lower case letters or spaces, and entity boundaries are obvious. But Chinese text is not clearly delineated, and entity boundaries are not clearly delineated. Text divided by commas and full stops suffers from problems such as excessive text length and too many entities. So advanced word segmentation techniques are needed for text segmentation. In addition, with the development of online language, the Chinese dataset is updated slowly, further increasing the difficulty of Chinese text entity recognition.

1.1. Related Work.

There are many research methods for named entity recognition, classified by category into dictionary-based, machine learning, deep learning, and pre-trained models. The dictionary-based approach involves string matching of text with a dictionary containing all entity names according to a specific policy or automatic recognition of text by domain experts who manually and extensively define the rules of linguistics. Quimbaya et al. [3] used a manual rule-based dictionary to extract text entities of electronic health records. Humphreys et al. [4] developed the LaSIE-II system for the MUC-7 task. Collins and Singer [5] proposed the impact of untagged examples on the classification of named entities and the need to develop many rules to improve the efficiency of entity recognition. Aone et al. [6] proposed Facile. Black et al. [7] proposed SAR. Zhang and Wang [8] conducted an experimental study on the recognition and analysis of Chinese institutional names, mainly using artificial rules for high school names. Rule-based approaches tend to achieve high accuracy rates when the rules developed to reflect the properties of the text more accurately, but this relies heavily on the expertise of linguists, and the limited number of rules makes it difficult to identify an infinitely variable set of entities more comprehensively [9]. In the new dataset, the old lexical rules cannot be used because of changes in the dataset, requiring rewriting of rules, building the knowledge base, and reconsidering the system's recognition capabilities, at great cost to these tedious efforts. At the end of the 20th century, machine learning algorithms were applied to named entity recognition tasks. The main methodological models are Conditional Random Field (CRF), Hidden Markov Model (HMM) and Decision Trees (DT) [10]. Statistical machine learning algorithms learn knowledge from a large annotated dataset without the need for manually defined rules. Daniel et al. [11] proposed an HMM-based textual entity recognition model that can effectively recognize and extract entities such as names, dates, times, and numbers. Le and Zhao [12] used HMM algorithm to identify named entities of Beijing Opera institutions. Bender et al. [13] proposed a framework for extracting named entities using a maximum entropy algorithm for natural language input. Ju et al. [14] used the SVM algorithm to identify biomedical named entities with good results. Shan et al. [15] proposed a method for relationship extraction of military named entities combining word rules and SVM models. Wang et al. [16] used conditional random fields as the basic framework for entity recognition of Uyghur texts. Lee et al. [17] used a conditional random field to develop a biomedical text recognition. The machine learning approach is more effective than the previous rule-based and lexicon-based approaches, eliminating the need for manual rule-making, reducing some of the costs, and making the constructed model portable and robust. However, the performance of statistical machine learning algorithms relies heavily on the features used and therefore requires extensive manual annotation by people with specialist domain knowledge. Deep learning neural network methods have become a major trend in natural language processing research. The current research methods for deep learning techniques in named entity recognition include Convolutional neural networks (CNN), Recurrent neural networks (RNN), Gated Recurrent Units (GRU), Bidirectional Long Short Term Memory (BiLSTM) networks, and so on [18-21]. Hammeerton [22] was the first to use the Long Short Term Memory (LSTM) network structure for textual entity recognition. The network is optimized based on the RNN network structure and is suitable for the classification, processing, and prediction of time series based data. The model is also validated on English and German development data. LSTM-CRF structure gradually became the basic structure of entity recognition. Gao et al. [23] proposed an LSTM-CRF model using CRF combined with LSTM for entity recognition of military movement text. Later, Lample et al. [24] proposed a model combining a bi-directional LSTM network and CRF that can acquire textual bi-directional semantic information and perform well in the task of textual named entities. He et al. [25] compared the performance of word-level statistical based methods and found that word-level named entity recognition was better for word-level named entities. Some researchers have also tried incorporating word-level features into the word-level named entity recognition structure for comprehensive training. Zhu [26] proposed a method for Uyghur text entity recognition to address the problems of lack of semantic information and its sparse data in the Uyghur text. Cui et al. [27] proposed an ancient Chinese named entity recognition algorithm based on the Lattice LSTM model. Ren et al. [28] added lexical and domain feature graph convolutional networks to the character representations obtained from BERT to effectively capture the constraint relations of distant words in sentences. Shen et al. [29] used a joint model of Star Transformer and TextCNN to extract sentence features. Wang et al. [30] proposes a multilayer dense attention model for image caption. Zhao et al. [31] investigated a Chinese attraction entity recognition model incorporating language models for attraction aliasing in tourist travel text entity recognition. Yue et al. [32] proposed a model based on BERT-BiLSTM-Attention-CRF for identifying and extracting relevant named entities. The corresponding entity tagging note specification was designed for different cases. Chen and Xia [33] proposed a framework combining instance migration and model migration for cross-domain named entity recognition of questioned texts of liver cancer patients with only a small number of annotations. The above methods have achieved some results in NLP and text entity recognition. However, all the above methods cannot deal with the problem of multiple-word meanings and can only deal with the feature vectors of independent characters and words. Google team combined the advantages of different language models and proposed the BERT model in 2018 [34].

1.2. **Motivation and contribution.** We proposed a new structure for Chinese text entity recognition. The model first pre-trains the word vectors using BERT, and inputs the obtained word vector information to the BiLSTM layer. The multi-headed attention module is introduced to enable the assignment of different weights according to the importance of words to the classification results and enhance the word dependency in the text. Finally, the best sequence is predicted by CRF decoding and the entity recognition results are output. We used MSRA and People's Daily datasets, and the experimental results show that excellent results are achieved on the two datasets with F_1 values of 96.14% and 96.37%, respectively. There are three main innovations in the paper: 1) The BERT pre-training model is applied to the Chinese entity recognition task, and the trained results are directly used as the BiLSTM-CRF layer input, reducing the model training workload and improving the model speed. 2) The dynamic word vector output from the BERT pre-training model characterizes more semantic information. In addition, full word MASK, an improved version of the original MASK, is used to predict the whole word, enabling the BERT model to learn word boundaries and better characterized. 3) Based on the BiLSTM model, the strategy of the feature extraction layer is improved by introducing a multi-headed self-attention mechanism. It can prevent the BiLSTM from losing much information, which is extremely important for text entity recognition. They are important to the classification results and capture multiple semantics from the character, word, and sentence levels.

2. BERT-BiLSTM-MHA-CRF model.

2.1. **Overview of the model.** The overall architecture of the model in this paper is shown in Figure 1. The model consists of several modules, including the Embedding embedded layer, MHA-BiLSTM layer, and CRF layer. The function of the embedding layer is to transform the original input text sequence into a model-acceptable digital vector matrix and further map it into the corresponding low-dimensional embedding vector representation. The embedding vectors output from the embedding layer are then fed into the multi-headed self-attentive module MHA and BiLSTM network module, respectively. The BiLSTM network module is used to learn the timing features and context information of the text sequence, while the multi-head self-attention mechanism is used to obtain the global feature representation of the input text sequence and the correlation strength between various characters. Finally, the outputs of the above two neural network modules will be spliced together and input to the CRF layer to obtain the final tag sequence.

The entity recognition algorithm pseudo code is shown in Algorithm 1.

2.2. **BERT model.** The word vectors trained by language models such as Glove, Word2Vec, and GPT are static vectors and cannot distinguish between different meanings of the same words. For example, the word "long" in "grow long fast" and "the road is 3 meters long" indicate a different meaning, but in the traditional word vector language model, the two "long" vectors have the same value. The GPT language model is a one-way model, which can represent multiple meanings of a word but cannot obtain information about the preceding and following words. The BERT model no longer uses the static text representation method word2vec, combined with the bidirectional encoder Transformer to

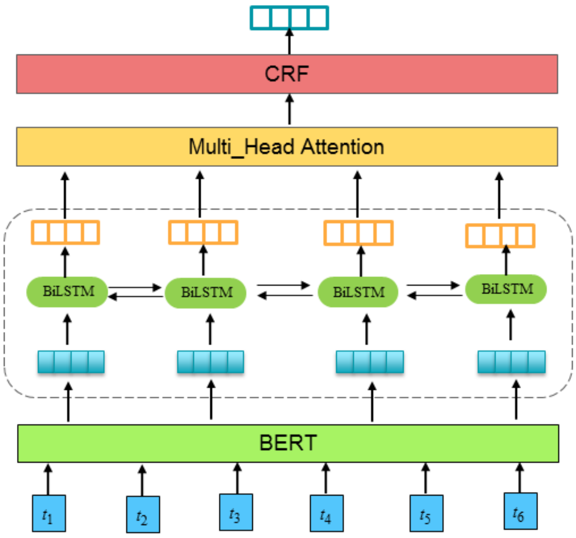

Figure 1. The proposed network structure

Algorithm 1 Entity recognition algorithm pseudo-code

Input: Entity recognition training set
Output: Entity identification results
1: Import BERT pre-training model
2: Set the maximum sequence length, model training parameters, etc.
3: Connect to BiLSTM layer
4: Set input gate, forget gate, output gate and other parameters
5: Connect to MHA layer
6: Set Q, K, V and other parameters
7: Connect to CRF layer
8: Train the entity recognition model
9: return entity recognition result

embed the text with words. The BERT model has a strong semantic acquisition capability to improve the recognition and extraction of entities and entity relationships, and the model is pre-trained using the MASK language model to predict the next word in a manner similar to completing a fill-in-the-blank. The traditional language model predicts the next word based on each given the word in the sentence, while the MASK model in the sentence and predicts the masked words of the contextual content. To perform Chinese text entity recognition better, the BERT model adopted the full word MASK method. The traditional BERT model slices the text in characters, dividing a complete word into several subwords, and these subwords will be randomly MASK during the model training. Full-word MASK requires splitting the Chinese dataset into words and then MASK each token belonging to the same word. In this way, the pre-trained model does not predict the individual token that is MASK, but each token within the same word that is MASK, is shown in Table 1.

Each input word vector is trained by the BERT model, as shown in Figure 2.

We found that the BERT pre-training model included the word, sentence, and position vectors. The word vector directly represents the information of the word itself; the BERT model automatically learns the sentence vector to represent the global semantic

Table 1. Full word MASK

Original text	Entity recognition is the key
Text to be split	Entity recognition is the key
Original MASK input	[MASK]recognition is the key
Full word MASK input	[MASK][MASK][MASK]is the key

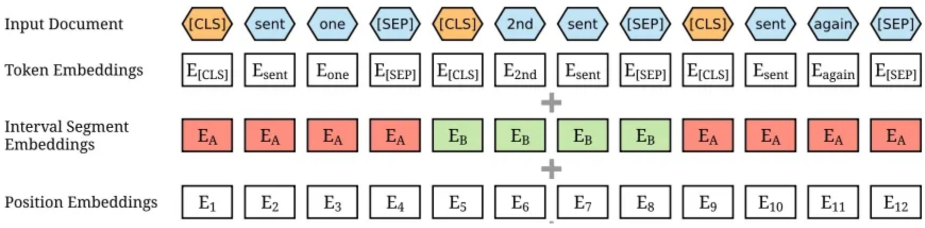

Figure 2. BERT structure diagram [34]

information of the word; the location vector information is used to distinguish the semantic information represented by the word in different positions, which the location vector can distinguish. The model's output is the fusion of the three vector information of the input words as the final output information. The [CLS] indicated the beginning of a text sequence, and the [SEP] indicated the interval between sentences or the end of a text sequence.

2.3. **BiLSTM model.** The LSTM structure enabled the RNN network to have a specific memory capability, which can effectively handle the task of text entity recognition. However, as the length of the processed text sequence increases, the RNN network may suffer from gradient explosion, affecting the model's operational effect. LSTM can use the gating mechanism to achieve the long-term memory function of the network and capture the above sequence information. The LSTM unit structure consists of three parts: input gate, forgetting gate and output gate, and the structure is shown in Figure 3. The three gate nodes can effectively overcome the problems of gradient explosion and gradient disappearance in RNN networks.

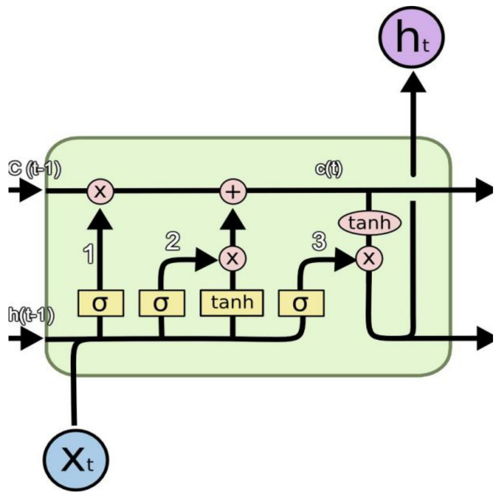

Figure 3. LSTM structure [22]

2.4. **Attention mechanisms.** The BiLSTM model cannot adequately represent the global information of the text sequence and the weight information of each character, and there is a risk of losing the information related to named entities as the sentence length increases. The attention mechanism completes the attention computation in the text series and finds the internal connection in the text sequence. Each word in the text is used as a query, key, and value at the same time, and each word is compared with other words. The relationship between two words is calculated, thus normalized to get the weight, and finally the weighted average of the word vectors of all words in the whole sentence is used as the new word vector of this word, which is traversed once to complete the update of the word vector in the sentence. The attention mechanism model overcomes the distance dependence problem of the BiLSTM layer to some extent.

2.5. **CRF layer.** The BiLSTM model can effectively handle long-range text information, model the contextual information in the input sequence, calculate the specific score of each label, and select the maximum score as the output label. However, the BiLSTM model cannot handle the dependency of adjacent labels. For example, the "I-ORG" label cannot be immediately followed by the "B-PER" label, resulting in the output labels cannot form a complete entity and cannot be used as the prediction result of the model. Therefore, we finally use the CRF layer to decode the new fused features produced by the BiLSTM-MHA output layer to find the global optimal character label sequence.

The CRF layer uses the output sequence of the upper layer $Z = \left[st_1, st_2, st_3,\dots, st_n\right]$ as input information to predict the most probable character tag sequence $Y = \left[y_1, y_2, y_3,\dots, y_n\right]$ based on the post contextual character tags. The score of the predicted sequence Y is obtained by summing the transfer probability matrix A with BiLSTM-MHA output layer Z according to Equation (1). Then the corresponding conditional probability p is obtained by the normalization operation.

$$Score_\theta(Z, Y) = \sum_{t=1}^{n}(Z_{y_t,t} + A_{y_{t-1},y_t}) \tag{1}$$

θ denote the set of parameters of the CRF layer. We could obtain the estimates of all parameters as shown in Equation (2).

$$L(\theta) = \sum_{(X,Y)\in y^X} logp(Y \mid Z, \theta) \tag{2}$$

3. Analysis of experimental results.

3.1. **Experimental data and performance measures.** The datasets used for the experiments are publicly evaluated Chinese datasets in China, namely the MSRA dataset and the People's Daily dataset. Three entities are included in the two datasets: Person Name, Location, and Organization. The specific annotation examples are shown in Table 2.

To verify the effectiveness of the proposed integrated model, we conducted experiments on the People's Daily dataset and the MSRA dataset. The dataset was divided into 60% as the training set, 20% as the validation set, and 20% as the test set, as shown in Table 3.

In the paper, the $Precision$, $Recall$, and F_1 values are used to measure the effectiveness of the model, in which the accuracy refers to the ratio of the number of correct entities to the total number of entities identified. The recall rate refers to the ratio of the number of entities correctly identified to the total number of entities. In some cases, there will be a conflict between accuracy and recall rate, so the F_1 value is used to consider P and R comprehensively. The calculation formula for each index is as follows:

Table 2. Examples of markings

Chinese	English	Tags
guo	Defence	B-ORG
fang	Minister	I-ORG
bu	General	I-ORG
zhang	chi	O
chi	hao	B-PER
hao	tian	I-PER
tian		I-PER
shang		O
jiang		O

Table 3. Dataset size statistics

Dataset	Training	Verification	Test
People's Daily	801672	267314	267080
MSRA	1435571	478563	478480

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{3}$$

$$R = \frac{T_p}{T_p + F_n} \times 100\% \tag{4}$$

$$F_1 = \frac{2PR}{P + R} \times 100\% \tag{5}$$

In the formula, T_p, F_p, and F_n are the number of real cases, false positive cases, and false counterexamples, respectively.

3.2. **Experimental environment and parameters.** The deep learning framework for model validation used in Tensorflow and the detailed experimental parameters are shown in Table 4.

Table 4. Experimental environment and parameters

Item	Environment
Operating systems	Windows 7 Signature Edition
Memory	12G
CPU	Intel i5-10400 2.9GHZ
Hard Disk	1T
Python	3.6.8
TensorFlow	1.14.6

We used both the original MASK and full-word MASK, respectively. During the training process, the adaptive moment optimization algorithm is used. The parameter settings are shown in Table 5.

Table 5. Training model parameters

Parameter name	Parameter values
Number of Transformer layers	12
Number of hidden layers	768
max_seq_len	100
Optimizers	Adam
Dropout	0.5
Learning Rate	0.001
Batch size	64

3.3. **Experimental results.** The Chinese entity models proposed in this paper are compared with other entity recognition models. To conduct a more objective evaluation, we evaluated the MSRA dataset and the People's Daily dataset, respectively, and the specific evaluation results are shown in Tables 6 to 7.

Table 6. Results of the People's Daily dataset

Models	P	R	F_1
LSTM-CRF	0.8299	0.8909	0.8574
BiLSTM	0.8392	0.8915	0.8639
BiLSTM-CRF	0.8711	0.8980	0.8840
IDCNN-CRF	0.8301	0.9128	0.8648
BERT-BiLSTM–CRF (original MASK)	0.9580	0.9364	0.9388
BERT-BiLSTM-CRF (full word MASK)	0.9611	0.9403	0.9492
BERT-BiLSTM-MHA-CRF (full word MASK)	0.9721	0.9595	0.9637

Table 7. Results of the MSRA dataset

Models	P	R	F_1
LSTM-CRF	0.8549	0.8622	0.8584
BiLSTM	0.8281	0.8768	0.8509
BiLSTM-CRF	0.8540	0.8876	0.8698
IDCNN-CRF	0.8301	0.9128	0.8648
BERT-BiLSTM–CRF (original MASK)	0.9637	0.9528	0.9489
BERT-BiLSTM-CRF (full word MASK)	0.9649	0.9471	0.9504
BERT-BiLSTM-MHA-CRF (full word MASK)	0.9668	0.9582	0.9614

Tables 6 and 7 show that the F_1 values of the BiLSTM-CRF model are 2.01% and 1.89% higher than the LSTM-CRF model for the People's Daily dataset and the MSRA dataset, respectively. It indicates that the BiLSTM model acquires forward and backward text information, and the text entity recognition is better than the LSTM model. The result showed that the CRF module could effectively improve the model recognition effect. The F_1 values of the IDCNN-CRF and BiLSTM-CRF models on the People's Daily dataset were 86.48% and 88.40%, respectively; the F_1 values on the MSRA dataset were 86.48% and 86.98%, respectively, indicating that the recognition results of these two models are relatively close. The word vectors pre-processed by BERT was used as input information for the BiLSTM model, and the experimental results showed a high improvement. F_1 values reached 96.37% and 96.14% in the People's Daily and MSRA datasets. Compared with the BiLSTM-CRF model, the improvement was 7.97% and 9.16%, respectively. Because the BERT model has strong semantic acquisition capability and can fully characterize information at the character level, word level, sentence level, and relations between utterances, after pre-processing by the BERT model, the trained word vectors can handle syntactic and word information in different contexts, enhancing the generalisation capability of the model and thus improving the ability of entity and entity relation recognition and extraction. The multi-headed self-attention mechanism assigns different weights to the words according to their importance to enhance the word dependencies within the text. To reduce the limitations of the model in semantic recognition, feature extraction is performed separately in the subspace of multiple feature representations. Then the results are fused to improve the learning ability of the model. Compared with BERT-BiLSTM-CRF, the F_1 values in the People's Daily and MSRA datasets are improved by 1.45% and 1.1%, respectively, after adding the multi-headed attention mechanism without introducing MHA. The full word MASK is an improved version of the original MASK, enabling the BERT model to learn word boundaries and better characterize the whole word. Compared with the original MASK, the F_1 values of the full word MASK model in the People's Daily dataset and the MSRA dataset are 1.04% and 0.15% higher than the original MASK model, respectively.

3.4. **Comparison of related models.** The results of the comparative analysis with the published models are shown in Tables 8 and 9.

Table 8. Comparison of models in the MSRA dataset

Models	P	R	F_1
LAC-DGLU [35]	0.9323	0.9358	0.9340
BiLSTM-Gate-NER [36]	0.9333	0.9215	0.9274
Bi-CRNN-GRU [37]	0.9362	0.9249	0.9305
IDC-HSAN [38]	0.9360	0.9265	0.9308
DNN-Fusion-Model [39]	0.9458	0.9447	0.9452
BERT-DNN-CRF [40]	0.9471	0.9421	0.9446
Transformer-Bert [41]	0.9607	0.9512	0.9559
WSA-CNER [42]	0.9437	0.9270	0.9366
BERT-BiLSTM-MHA-CRF (full word MASK)	0.9668	0.9582	0.9614

Table 9. Comparison of test models in the People's Daily dataset

Models	P	R	F_1
BiLSTM-Gate-NER [36]	0.9226	0.9058	0.9141
Bi-CRNN-GRU [37]	0.9468	0.9382	0.9425
BERT-DNN-CRF [40]	9454	0.9410	0.9432
Transformer-Bert [38]	0.9545	0.9130	0.9334
CCRFs [43]	0.8812	0.9005	0.8907
Deep-Belief-Net [44]	0.8923	0.8994	0.8958
BiLSTM [45]	0.9571	0.9359	0.9183
BERT-BiLSTM-MHA-CRF (full word MASK)	0.9721	0.9595	0.9637

From Tables 8 and 9, we found that the LAC-DGLU model used the word embedding algorithm of Local Attention Convolution and a DGLU structure modified based on CNN to alleviate the model's reliance on the word separation effect. The BiLSTM-Gate-NER model proposed a cascading deep neural network model that parses the entity recognition of each layer into a separate task and then facilitates the information exchange between layers through the Gate filtering mechanism. The Bi-CRNN-GRU model relies on a convolutional recurrent neural network to build a feature encoding layer to realize the joint extraction of local spatial features and long-range time-dependent features of Chinese text sequences. The IDC-HSAN model is a high-performance parallelizable NER model with multiple granularities as input and a hierarchical attention mechanism; the CCRFs model proposes a new automatic recognition algorithm for Chinese institutional names based on the CRF, and the recognition algorithms are designed separately according to the complexity of entity recognition. The BiLSTM model uses context-based word vectors and word-based word vectors to constrain the cost function of the BiLSTM using the correlation between labels in the annotation sequence and embedding domain knowledge into the cost function of the model to enhance the recognition capability of the model further. However, the various above improved models can learn feature information of characters or words, and the model performance improvement is limited. The highest F_1 value achieved was 88.40% in the People's Daily dataset and 86.98% in the MSRA dataset. Through comparative experimental analysis with other models, we found that the BERT-BiLSTM-MHA-CRF model has the best performance in all aspects among all the compared models, indicating the superiority of the BERT pre-training model in the Chinese text recognition process. The entity recognition model proposed in this paper has made some achievements, the F_1 value on the two datasets is more than 96%, but it cannot reach 100%. The reason is that the Chinese text has specific structures, including subject-verb-object-definitive-complement, and it is difficult to identify and distinguish entities.. At the same time, there are polysemous words in Chinese, which will interfere with entity recognition. For example, "I bought Xiaomi on Xiaomi cell phone", the word "Xiaomi" is difficult to distinguish whether it means Xiaomi Technology Company or Xiaomi (grain). These multiple meaning words cannot be avoided in the process of text entity recognition, and word sense disambiguation techniques can be used to improve entity recognition.

4. **CONCLUSION.** Natural language processing research has been one of the hot spots of deep learning research in recent years. We proposed an improved text entity model structure based on the BERT model. The BERT pre-processing model processes the input Chinese text vectors. To improve the model adaptability, two kinds of BERT pre-training models, the original MASK and the full-word MASK, are used in this paper for comparison experiments. Then, the BERT pre-trained word vectors are used as the input information of BiLSTM for bi-directional information processing. Finally the optimal label sequences are obtained by CRF decoding to improve the text entity recognition effect. The F_1 values in the MSRA and People's Daily datasets reached 96.14% and 96.37%, respectively. Compared with the classical BiLSTM-CRF model, the improvements are 7.97% and 9.16%, respectively. Compared with all published models, the improved model proposed also achieved very good results. The method proposed in this paper also has some limitations. The next step is to simplify the model structure and improve the model training speed on the one hand and apply the model to other datasets to improve the model adaptability on the other hand.

Acknowledgment. This work was supported in part by the Excellent Young Talents in Anhui Universities Project (Granted No. gxyq2022026), Anhui Province Quality Engineering Project (Granted No. 2021jyxm0801), Teaching Research Key Project of Anhui

University of Chinese Medicine (Granted No. 2020xjjy_zd001) and Key Projects of Humanities and Social Sciences in Anhui Province (Granted No. 2022AH050428).

REFERENCES

[1] B. Sundheim, "Named entity task definition," in *Sixth Message Understanding Conf (MUC-6)*. 1995, pp. 319-332.

[2] M.-S. Sun, C.-N. Huang, H.-Y. Gao, and J. Fang, "Identifying Chinese names in unrestricted texts," *Communication of COLIPS,* vol. 4, no. 2, pp. 113-122, 1994.

[3] A.-P. Quimbaya, A.-S. Múnera, R.-A. Rivera, J.-C. Rodríguez, O.-M. Velandia, A.-A. Peña, and C. Labbé, "Named entity recognition over electronic health records through a combined dictionary-based approach," *Procedia Computer Science*, vol. 100, pp. 55-61, 2016.

[4] K. Humphreys, R. Gaizauskas, S. Azzam, C. Huyck, B. Mitchell, H. Cunningham, and Y. Wilks, "University of Sheffield: Description of the LaSIE-II system as used for MUC-7," in *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference*. Virginia, 1998, pp. 1-20.

[5] M. Collins, and Y. Singer, "Unsupervised models for named entity classification," in *1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*. 1999, pp. 100-110.

[6] C. Aone, L. Tom Hampton, and M. Ramos-Santacruz, "SRA: Description of the IE2 system used for MUC-7," in *Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference*. Virginia, 1998, pp. 1-14.

[7] W.-J. Black, F. Rinaldi, and D. Mowatt, "FACILE: Description of the NE System Used for MUC-7," in "*Seventh Message Understanding Conference (MUC-7): Proceedings of a Conference*. Virginia, 1998, pp. 1-10.

[8] X.-H. Zhang, L.-L. Wang, "Identification and analysis of Chinese institutional names," *Journal of Chinese Information Processing*, vol. 11, no. 4, pp. 21-32, 1997.

[9] L. Chiticariu, R. Krishnamurthy, Y.-Y. Li, F. Reiss, and S. Vaithyanathan, "Domain adaptation of rule-based annotators for named-entity recognition tasks," in *Proceedings of the 2010 Conference*. 2010, pp. 1002-1012.

[10] F.-Q. Zheng, L. Lin, B.-Q. Liu, and C.-J. Sun, "A Research on the Application of HowNet in Named Entity Recognition," *Journal of Chinese Information Processing*, vol. 22, no. 05, pp. 97-101, 2008.

[11] D.-M. Bikel, R. Schwartz, and R.-M. Weischedel, "An algorithm that learns what's in a name," *Machine learning*, vol. 34, no. 1, pp. 211-231, 1999.

[12] J. Le, and X. Zhao, "HMM-based named entity recognition algorithm for Beijing opera institutions," *Computer Engineering*, vol. 39, no. 6, pp. 266-271, 2013.

[13] O. Bender, F.-J. Och, and H. Nez, "Maximum entropy models for named entity recognition," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 2003, pp. 148-151.

[14] Z. Ju, J. Wang, and F. Zhu, "Named entity recognition from biomedical text using SVM," in *5th International Conference on Bioinformatics and Biomedical Engineering*. IEEE, 2011, pp. 1-4.

[15] H.-Y. Shan, Z.-L. Wu, H.-S. Zhang, and P.-L. Liu, "A military named entity relationship extraction method combining word rules and SVM model," *Command Control and Simulation*, vol. 38, no. 4, pp. 58-63, 2016.

[16] L.-L. Wang, A.-S. Wumaier, M.-M. Maimaiti, K.-J. Abiderexiti, and T.-G. Yibulayin, "Named entity recognition in Uyghur based on CRF and semi-supervised learning," *Journal of Chinese Information*, vol. 32, no. 11, pp. 16-26, 2018.

[17] H.-C. Lee, Y.-Y. Hu, and H.-Y. Kao, "An enhanced CRF-based system for disease name entity recognition and normalization on BioCreative DNER Task," in *Proceedings of the Fifth Biocreative Challenge Evaluation Workshop*. 2015, pp. 226-233.

[18] F.-Z. Wu, J.-X. Liu, C.-H. Wu, Y.-F. Huang, and X. Xie, "Neural Chinese named entity recognition via CNN-LSTM-CRF and joint training with word segmentation," in *The World Wide Web Conference*. 2019, pp. 3342-3348.

[19] F. Yang, J.-H. Zhang, G.-S. Liu, J. Zhou, C. Zhou, and H.-R. Sun, "Five-stroke based CNN-BiRNN-CRF network for Chinese named entity recognition," in *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer, 2018, pp.184-195.

[20] Q. Wang, Y.-M. Zhou, T. Ruan, D.-Q. Gao, Y.-H. Xia, and P. He, "Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition," *Journal of Biomedical Informatics*, vol. 92, 103133, 2019.

[21] Y.-R. Ma, Y.-J. Peng, and T.-Y. Wu, "Transfer learning model for false positive reduction in lymph node detection via sparse coding and deep learning," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 2, pp. 2121-2133, 2022

[22] J. Hammerton, "Named entity recognition with long short-term memory," in *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*. 2003, pp. 172-175.

[23] X. Gao, J.-D. Zhang, X. Xu, and J.-H. Feng, "LSTM-CRF-based text entity recognition method for military movements," *Command Information System and Technology*, vol. 11, no. 6, pp. 91-95, 2020.

[24] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," 2016. [Online]. Available: https://doi.org/10.48550/arXiv.1603.01360.

[25] J.-Z. He, and H.-F. Wang, "Chinese named entity recognition and word segmentation based on character," in *Proceedings of the Sixth SIGHAN Workshop on Chinese Language Processing*. 2008, pp. 128-132.

[26] S.-L. Zhu, "A deep learning-based model for recognition of named entities in Uyghur," *Computer Engineering and Design*, vol.40, no. 10, pp. 2874-2878, 2019.

[27] D.-D. Cui, X.-L. Liu, R.-Y. Chen, X.-H. Chen, Z. Li, and L. Qi, "Lattice LSTM-based named entity recognition for ancient Chinese," *Computer Science*, vol. 47, no. S2, pp. 18-22, 2020.

[28] Q.-T. Ren, H. Wang, X. Xiong, and T. Fan, "Construction of a model for extracting non-heritage theatre terms with GCN teleconstraint and its application research," *Data Analysis and Knowledge Discovery*, vol. 5, no. 6, pp. 1-19, 2021.

[29] H. Shen, Y.-J. Zhang, B.-H. Xie, and H.-Y. Zhao, "Chinese named entity recognition based on BSTTC model," *Computer System Applications*, vol. 30, no. 6, pp. 262-270, 2121.

[30] E.-K. Wang, X. Zhang, F. Wang, T.-Y. Wu, and C.-M. Chen, "Multilayer dense attention model for image caption," *IEEE Access*, vol. 7, pp. 66358-66368, 2019.

[31] P. Zhao, L.-Y. Sun, Y. Wan, and N. Ge, "Named entity recognition of Chinese attractions based on BERT-BiLSTM-CRF," *Computer System Applications*, vol. 29, no. 6, pp. 169-174, 2020.

[32] Y. Wang, M.-X. Wang, S. Zhang, and Z.-B .Du, "BERT-based named entity recognition for police text," *Computer Applications*, vol. 40, no. 2, pp. 535-540, 2020.

[33] M.-S. Chen, and C.-X. Xia, "A study of named entity recognition for online questions from liver cancer patients:a migration learning-based approach," *Data Analysis and Knowledge Discovery*, vol. 3, no. 12, pp. 61-69, 2019.

[34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018. [Online]. Available: https://doi.org/10.48550/arXiv.1810.04805.

[35] F. Zhao, J. Huang, and C. Zhang, "LAC-DGLU:A named entity recognition model based on CNN and attention mechanism," *Computer Science*, vol. 47, no. 11, pp. 212-219, 2020.

[36] Y.-L. Jin, J.-F. Xie, and D.-J. Wu, "Chinese nested named entity recognition based on hierarchical annotation," *Journal of Shanghai University (Natural Science Edition)*, vol. 28, no. 2, pp. 270-280, 2021.

[37] X.-H. Huang, L.-X. Qiao, W.-T. Yu, J. Li, and H. Xue, "Joint learning of Chinese word separation and named entity recognition," *Journal of the National University of Defense Technology*, vol. 43, no. 1, pp. 86-94, 2021

[38] R. Chen, and X.-L. Lu, "Fusion of null convolutional neural network and hierarchical attention mechanism for Chinese named entity recognition," *Journal of Chinese Information*, vol. 34, no. 8, pp. 70-77, 2020.

[39] D.-D. Zhao, J.-X. Cao, D.-G. Huang, J.-N. Meng, and P. Zhang Dual, "Neural Network Fusion Model for Chinese Named Entity Recognition," *International Journal of Computational Intelligence Systems*, vol. 14, no. 1, pp. 471-481, 2021.

[40] Q. Cai, "A multi-criteria fusion approach for Chinese named entity recognition," *Journal of Southeast University (Natural Science Edition)*, vol. 50, no. 5, pp. 929-934, 2020.

[41] Y.-Y. Liu, Z.-T. Yu, S.-X. Gao, J.-J. Guo, Y.-F. Zhang, and B.-J. Nie, "A Chinese named entity recognit-ion method based on machine reading comprehension," *Pattern Recognition and Artificial Intelligence*, vol. 33, no. 7, pp. 653-659, 2020.

[42] S.-S. Zhong, X. Chen, M.-H. Zhao, and Y.-J. Zhang, "A Chinese named entity recognition method introducing a word set-level attention mechanism," *Journal of Jilin University (Engineering Edition)*, vol. 52, no. 5, pp. 1098-1105, 2021.

[43] J.-S. Zhou, X.-Y. Dai, C.-Y. Yin, and J.-J. Chen, "Automatic recognition of Chinese organization names based on cascaded conditional random field model," *Journal of Electronics*, no. 5, pp. 804-809, 2006.

[44] D. Yin, J.-S. Zhou, and W.-G. Qu, "Chinese nested named entity recognition based on joint model," *Journal of Nanjing Normal University (Natural Science Edition)*, vol. 37, no. 3, pp. 29-35, 2014.

[45] Y.-H. Feng, H. Yu, G. Sun, and J.-J. Sun, "Named entity recognition method based on BLSTM," *Computer Science*, vol. 45, no. 2, pp. 261-268, 2018