Journal of Network Intelligence ©2022 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **7**, Number **2**, May **2022**

Medical Text Entity Study based on BERT-BiLSTM-MHA-CRF Model

Tongping Shen*
[1] School of Medicine and Information Engineering
Anhui University of Chinese Medicine
Hefei, 230012, China
[2] Graduate school, Angeles University Foundation,
Angeles City, Philippines
Corresponding Author: shentp2010@ahtcm.edu.cn

Huanqing Xu
School of Medicine and Information Engineering
Anhui University of Chinese Medicine
Hefei, 230012, China
huanqingxu@tju.edu.cn

Received February 2022; revised March 2022

Abstract. *In the process of natural language processing research, text entity recognition research has been the key research direction. Traditional language processing models cannot effectively represent the contextual semantic information in text and cannot handle different contexts with multiple meanings of a word, which affects the effect of entity recognition. In the paper, we propose a full-word-based text entity recognition model ,just that a BERT-BiLSTM-MHA-CRF model, in which the BERT pre-processing language model generates word vectors that represent contextual semantic information, and the whole-word MASK processing mechanism is more suitable for processing Chinese text. The BERT preprocessing layer is able to generate word vectors characterizing contextual semantic information, the BiLSTM layer extracts bidirectional feature information from the input word vectors, the attention mechanism layer assigns weights to the output feature vectors to effectively obtain long-range dependencies in text utterances, and finally decodes them by CRF to generate entity label sequences. The experimental results show that the model achieves excellent performance on both the Resume corpus and the CCKS2017 corpus, with F1 reaching 96.14% and 92.68% respectively. The medical text entity recognition model proposed in this paper can further improve the recognition effect and accuracy rate, which has a positive impact on the subsequent paperless application of electronic medical records and automatic diagnosis of medical images.*
Keywords: Chinese electronic medical record, named entity recognition, deep learning, multi-head attention, BERT

1. **Introduction.** The study of text named entity recognition has developed rapidly and received widespread attention, the research mainly focuses on in finance, news media, medical text processing and so on. Text entity recognition is an important foundation for research on information extraction, intelligent question and answer, and knowledge graph, etc. The merit of NER extraction directly affects the effect of subsequent natural processing [1]. Chinese named entities are more difficult to identify than English. 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 corpus is updated slowly, further increasing the difficulty of Chinese text entity recognition.

With the acceleration of the national medical informatization process, the volume of electronic medical record data is rapidly increasing, and electronic medical records contain a large amount of medical tacit knowledge. Relevant studies show that electronic medical records are knowledge intensive texts with a higher density of medical entity distribution than general purpose domain texts, which have very important research value. Compared with general domain texts, the entity types of texts in electronic medical records mainly include symptoms, disease names, examination means and modalities, etc. The large number and rich types of these entities, as well as the variation of entity length and the existence of aliases and acronyms in entity structure, cause the poor recognition of entities in electronic medical record texts.

In recent years, traditional lexical methods, machine learning and deep learning have been widely used for entity recognition tasks.

Subsequently traditional machine learning methods were applied in entity recognition models. Nan Yu et al. used a multi-feature fusion CRF model for medical entity recognition [2]. Li et al. used HMM algorithm for text sentiment analysis [3]. In traditional machine learning approaches, named entity recognition requires a large scale corpus to learn the annotation model, and requires human involvement in feature extraction, and is highly dependent on the quality of corpus annotation.

Deep learning techniques are also gradually being applied to entity recognition research tasks. Lian used lexicon construction, word vector training, sequence annotation, and model training methods to build Bi-LSTM-CRF models to identify 12 classes of named entities in the cyberspace security domain [4]. Zhu proposed a deep neural network-based method for Uyghur named entity recognition to address the problems of lack of semantic information and its sparse data in Uyghur named entity recognition [5].

The above methods have achieved some success in the field of natural language processing, and the efficiency of entity recognition has been further improved. However, all the above methods have a problem that they cannot deal with the problem of multiple meanings of words, and can only deal with independent character and word feature vectors, ignoring the contextual semantic information of characters, leading to the problem of poor entity recognition accuracy. The BERT model uses a bidirectional transformer neural network as an encoder to enhance the generalization ability of the pre-trained word vector model to fully characterize the relationship between characters, words and utterances. The prediction of the next word can refer to the input information in both front back direction, characterize the semantics of the same word in different contexts, and effectively solve the problem of multiple meanings of a word.

Based on the BERT-BiLSTM-CRF model, Ren et al. added lexical and domain features to the character representation obtained by BERT, followed by adding graph convolutional networks after BiLSTM to better capture the constraint relations of distant words in sentences [6]. Shen et al. proposed a Chinese named entity recognition method based on BSTTC model [7]. Yue et al. combined deep learning with knowledge graph and proposed a model based on improved BERT and bidirectional RNN for forestry entity recognition and entity relationship extraction [8]. Jiang et al. performed word vector pre-training through BERT layer to obtain entities in food case dispute adjudication documents[9]. In order to solve the problem of multiple meanings of a word in the feature representation of tourism text, Zhao et al. studied a Chinese attraction entity recognition model incorporating language models for attraction aliases in the recognition of attraction entities in tourism travel text[10].

Wang proposed BERT-BiLSTM-Attention-CRF for identifying and extracting relevant named entities for the problem of difficult identification of key entity information in the police domain [11]. Based on the BiLSTM+CRF model, Lian et al. fused the BERT pre-trained language model and neural network to extract cyberspace security entities [12].

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 input of the BiLSTM-CRF layer, which reduces the model training workload and improves the model speed; (2) the dynamic word vector output from the Bert pre-training model characterizes more semantic information. In addition, whole-word MASK is used, which is an improved version of the original MASK that can predict the whole word, enabling the BERT model to learn the word boundaries and better characterize the semantic information of the whole word. (3) In order to obtain more semantic information, a multi-headed attention layer is added between the BiLSTM layer and the CRF layer, which can effectively obtain the long-range dependencies in text utterances and improve the model recognition effect.

2. BERT-BiLSTM-MHA-CRF model.

2.1. **Overview of the model.** The BERT layer pre-trains the input text to generate dynamic word vectors, and uses the obtained word vector information as input to the BiLSTM layer for bi-directional training to further extract text features. The attention mechanism extracts the key features for entity recognition from the output of the BiLSTM layer, and assigns weights to the feature vectors of the upper layer, highlighting the features that are key for entity recognition and ignoring irrelevant features.

Through weight checking, it is straightforward to evaluate which embeddings are the preferred embeddings for a particular downstream task. Finally the CRF layer allows the dependencies between predicted labels to be effectively constrained and the sequence of labels to be modelled to obtain the globally optimal sequence. The proposed model architeture is shown in Figure 1.

2.2. **BERT model.** In natural language processing, converting textual information into corresponding word vectors for embedding into models is an important task in natural language processing. The word vectors trained by these models are static vectors and cannot solve the problem of multiple meanings of words. The GPT language model is a unidirectional model, which can represent multiple meanings of a word, but cannot capture the contextual information of the word.

The BERT model has a strong semantic acquisition capability to improve entity and entity relationship recognition and extraction, and uses the MASK language model to pre-train the model to predict the next word in a similar way to the completion of a fill-in-the-blank. The traditional language model predicts the next word based on each given word in the sentence, while the MASK language model masks 15% of the words in the sentence and predicts the masked words by the contextual content. In order to perform Chinese text entity recognition better, the BERT model in this paper adopts the full word MASK method. The traditional BERT model slices the text in characters, which will divide a complete word into several subwords, and these subwords will be randomly MASK during the model training. In full-word MASK, if the part of the word is MASK, the whole word will be MASK, which is more in line with Chinese text processing habits.

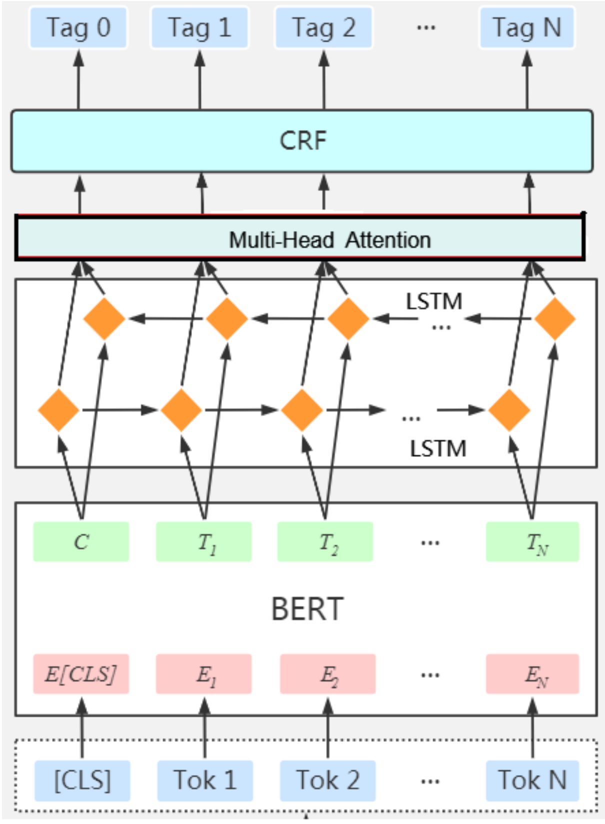

FIGURE 1. The proposed model architecture

In the process of word embedding, the pre-training results of BERT on massive text datasets are fully utilized, and the obtained word embedding incorporates contextual information, even if the same word is in different contexts, the obtained word embedding is different, which can dynamically represent the text containing different semantics in different contexts.

The word vector output after training with the BERT model consists of three components, a word vector, a sentence vector and a position vector. The input characters are converted into word vector form by looking up a word vector table. Sentence vectors represent semantic information in sentences and distinguish different utterances. The position vector distinguishes the semantic information of the words in different positions in different utterances. By pre-training the model with BERT, a text sequence vector containing rich semantic features can be obtained. Where the [CLS] special marker indicates the beginning of a text sequence and the [SEP] special marker indicates the interval between sentences or the end of a text sequence.

2.3. **BiLSTM module.** 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. The three gate nodes can effectively overcome the problems of gradient explosion and gradient disappearance in RNN networks.

One-way LSTM can only consider forward information in a text sequence and cannot handle backward information. In natural language processing, we slice the text information into individual words, and each word is related to the preceding and following words in some way, so the recognition of entities must consider the contextual relationship. To improve the adaptability of the model, the article uses the BiLSTM model, just as the

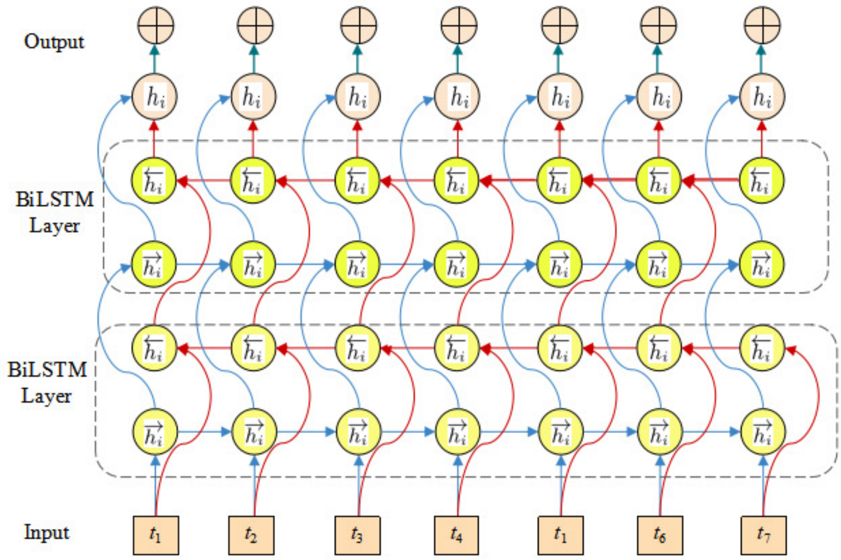

FIGURE 2. The BiLSTM network structure

input word vector and the output vectors are stitched together to calculate the output of the BiLSTM model. The BiLSTM network structure is shown in Figure 2.

2.4. **Attention mechanisms.** The attention mechanism is similar to the human vision mechanism in that when we observe an object, we generally focus on the important parts of the object and ignore the irrelevant detailed information. In the paper, the attention mechanism focuses on extracting feature information that plays a key role in entity recognition from the output of the BiLSTM layer, and assigning weights to the feature vectors output by the upper layer to highlight features[13]. The attention mechanism model can direct assessment of which embeddings are the preferred embeddings for a particular downstream task through weight checking, thereby improving the overall effectiveness of the model. For moment t the output of the model after weighting by the attention mechanism is shown in equation 1.

$$w_t = \sum_{n}^{j=1} \lambda_{tj} h_j \tag{1}$$

where w_t is the word feature vector weighted using the attention mechanism, h_j is the feature vector output by the BiLSTM model, n is the number of characters input to the model, and the weight λ_{tj} is calculated from the word feature vector w_{t-1} and h_j at moment $t-1$.

The self-attention mechanism completes the computation of attention in a text series, looking for internal connections in the text sequence. Each word in the text is used as query, key and value at the same time, and each word is compared with all the words in the sentence. The relationship between the two words is calculated and normalised to obtain the weights, and finally the weighted average of the word vectors of all words in the whole sentence is used as the new word vector for the word, which is traversed once to complete the update of the word vectors in the sentence. The Multi-Head attention structure is shown in Figure 3.

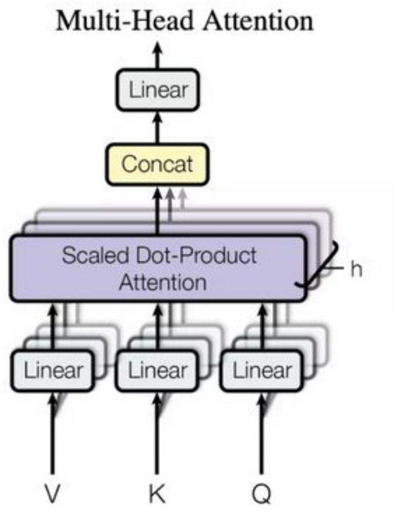

FIGURE 3. Multi-Head Attention

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_K}})V \tag{2}$$

Where Q, K, V denote the three matrices obtained after the same input is calculated with different parameters, usually $Q = K = V$. A moderating smoothing factor $\sqrt{d_K}$ for the k dimensions prevents the multiplication result from being too large. The $SoftMax()$ function normalises the result to a probability distribution and finally multiplies the matrix V to output the result[14].

In order to obtain contextual feature information in multiple dimensions, the paper adopts a multi-headed attention mechanism model, the main improvement lies in the fact that the query, key, and value are linearly transformed several times in different ways, and the parameters W are different for each linear transformation of Q, K, and V. In the paper, we used a multi-headed attention mechanism model, and the main improvement lies in the fact that the query, key, and value are linearly transformed several times in different ways instead of being computed only once, allowing the model to learn relevant information in different representation subspaces, which greatly improves the fitting ability of the model.

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{3}$$

$$Multi(Q, K, V) = Concat(head_i, \dots, head_h)W^o \tag{4}$$

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[15]. 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.

Using the output sequence of the upper layer $Z = \left[st_1, st_2, st_3, \dots, st_n\right]$ as input information, the CRF layer predicts the most probable character tag sequence $Y = \left[y_1, y_2, y_3, \dots, y_n\right]$ based on the pre and post contextual character tags.

The score of the predicted sequence Y is obtained by summing the transfer probability matrix A with the BiLSTM-MHA output layer Z according to equation 5, and then the corresponding conditional probability p is obtained by the normalization operation.

$$Score_\theta(Z, Y) = \sum_{t=1}^{n}(Z_{y_t,t} + A_{y_{t-1},y_t}) \tag{5}$$

$Z_{y_t,t}$ denotes the probability value of the input character w_t being tagged as a y_t tag at the current moment t,A_{y_{t-1},y_t} denotes the probability of the character at moment $t-1$ being tagged by w_t and being tagged as a y_t tag.

θ denote the set of parameters of the CRF layer, we can obtain the estimates of all parameters by maximizing the log-likelihood function, as shown in equation 6.

$$L(\theta) = \sum_{(X,Y)\in y^X} logp(Y \mid Z, \theta) \tag{6}$$

Y is the sequence of tags corresponding to the sequence of text characters, p denotes the conditional probability of Y given the input feature sequence Z and the set of parameters, and y^X is the set of all tags.

In the CRF model prediction process, the Viterbi algorithm is used to solve for the globally optimal sequence, with the formula shown in equation 7, where $y*$ is the sequence in the set that makes the score function achieve the maximum value.

$$y^* = arg \max_{\tilde{y}\in y^X} score(X, \tilde{y}) \tag{7}$$

3. Analysis of experimental results.

3.1. **Experimental data set.** In the paper, we use two publicly available datasets, such as the CCKS2017 electronic medical record dataset and the Resume dataset, which is mainly used for professional medical entity recognition from the perspective of electronic medical records. Therefore, the datasets used for model validation in this paper are scalable, with both conventional texts and professional medical texts. Table 1 and 2 lists the statistics for different categories of entities in the datasets.

Table 1. Statistics on different categories of entities in the dataset CCKS2017

Entity	Count	Train	Test
BODY	10179	8942	1777
CHECK	9546	7887	1659
DISEASE	722	515	207
SYMPTOM	7831	6477	1354
TREATMENT	1048	853	195

In the paper, we choose *Precision*, *Recall* and $F1$ value to evaluate the model, and the higher the evaluation index, the better the model performance.

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{8}$$

Table 2. Statistics on different categories of entities in the dataset Resume

Entity	Count	Train	Test
CONT	260	215	45
EDU	856	710	146
LOC	47	39	8
NAME	952	790	162
ORG	4611	3825	786
PRO	287	237	50
RACE	115	95	20
TITLE	6308	5233	1075

$$R = \frac{T_p}{T_p + F_n} \times 100\% \tag{9}$$

$$F_1 = \frac{2PR}{P + R} \times 100\% \tag{10}$$

3.2. **Experimental environment and parameters.** The entity recognition model experiments are based on the TensorFlow framework. In the paper, a BERT pre-trained language model is used, using both raw MASK and full-word MASK respectively. During the training process, the adaptive moment optimization algorithm is used.

3.3. **Experimental results.** The Chinese entity models proposed in this paper are compared with other entity recognition models respectively, and in order to make a more objective evaluation, the Chinese resume corpus and the CCKS2017 corpus are evaluated respectively, and the specific evaluation results are shown in Table 3 to 4.

Table 3. Results of the Resume Corpus Test (Unit: %)

Models	P	R	F_1
LSTM-CRF	0.8426	0.8761	0.8863
BiLSTM	0.8524	0.9019	0.8765
BiLSTM-CRF	0.8924	0.9148	0.9035
IDCNN-CRF	0.8957	0.9089	0.9024
BERT-BiLSTM–CRF (original MASK)	0.9392	0.9574	0.9482
BERT-BiLSTM-CRF (full word MASK)	0.9668	0.9582	0.9614

Table 4. Results of the CCKS2017 corpus test results (unit: %)

Models	P	R	F_1
LSTM-CRF	0.7968	0.8012	0.7939
BiLSTM	0.7931	0.8065	0.7845
BiLSTM-CRF	0.8083	0.8182	0.8132
IDCNN-CRF	0.8046	0.8244	0.8144
BERT-BiLSTM–CRF (original MASK)	0.9341	0.9234	0.9263
BERT-BiLSTM-CRF (full word MASK)	0.9333	0.9258	0.9268

Through Tables 3 and 4, we can found that the F1 values of BiLSTM-CRF model are 1.72% and 1.93% higher than those of LSTM-CRF model in resume corpus and CCKS2017 corpus respectively, indicating that BiLSTM model obtains forward and backward text information and the text entity recognition effect is better than that of LSTM model. And also , the BiLSTM-CRF model has a higher F1 value than that of BiLSTM model in Resume corpus and CCKS2017 corpus, indicating that the CRF module can effectively improve the model recognition effect. Because the CRF layer can effectively constrain the dependencies between the predicted labels and model the label sequences to obtain the global optimal sequences.

The F1 values of the IDCNN-CRF and BiLSTM-CRF models on the resume corpus are 90.24% and 90.35%, respectively; the F1 values on the CCKS2017 corpus are 81.44% and 81.32% , indicating that the recognition effects of these two models are relatively close.

The BERT are used as the input information of BiLSTM model, and the experimental results have high improvement. The F1 values in the resume corpus and CCKS2017 reach 94.82% and 92.63%, respectively. Compared with the BiLSTM-CRF model, the improvement is 4.47% and 11.31%, respectively. Because the Bert model has strong semantic acquisition ability and can fully characterize the information of word, the trained word vectors can handle syntactic and word information in different contexts after preprocessing by the BERT model.

The whole-word MASK is an improved version of the original MASK, which is able to predict the whole word, enabling the BERT model to learn word boundaries and better characterize the semantic information of the whole word. Compared with the original MASK, the F1 values of the whole-word MASK model in the resume corpus and the CCKS2017 corpus are 1.32% and 0.05% higher than the original MASK model, respectively, indicating that the performance of the original BERT model in the resume corpus and the CCKS2017 corpus is already very good, and the scope for improvement of the whole-word MASK model is limited.

3.4. **Comparison of related work.** Comparison experiments were designed for the resume corpus and the CCKS2017 corpus respectively, and the model validation results are shown in Tables 5 and 6.

Table 5. Comparison of test models in Resume corpus (unit: %)

Models	P	R	F_1
FLAT[16]	-	-	0.9545
CBiLSTM-CRF[17]	0.8943	0.9168	0.9054
SoftLexicon[18]	0.9530	0.9577	0.9553
Attention-BiLSTM-CRF[19]	0.8958	0.9206	0.9080
WSA-CNER[20]	0.9560	0.9522	0.9561
BERT-BiLSTM-CRF (full word MASK)	0.9668	0.9582	0.9614

From Tables 5 and 6, we can find that the core of the FALG model is to use the Chinese character lattice structure for entity recognition, and the lattice structure is converted into a set of spans and specific location encoding is introduced; CBiLSTM-CRF model proposes a method for unstructured text resume analysis based on character sequence model. The powerful learning ability of BLSTM is used to learn the features, extract the corresponding features, and use CRF to get the optimal label sequence according to the constraints of the before and after labels; SoftLexicon proposes a vector representation using lexicon information and adjusts the character representation layer sequence model

Table 6. Comparison of test models in CCKS2017 corpus (unit: %)

Models	P	R	F_1
Attention-BiLSTM-CRF[21]	0.9126	0.9038	0.9082
Dict-BiLSTM-CRF[22]	0.9083	0.9164	0.91244
BiLSTM-CRF[23]	0.9112	0.8947	0.9043
RNN-CRF[24]	0.9299	0.8925	0.9108
Attention-ID-CNNs-CRF[25]	0.9415	0.9463	0.9455
BERT-BiLSTM-CRF (full word MASK)	0.9333	0.9258	0.9268

of neural network for information modeling; Attention-BiLSTM-CRF proposes adding Attention mechanism and word association features to the BiLSTM neural network to build a BiLSTM deep learning model based on Attention mechanism; WSA-CNER model builds a word set-level Attention mechanism based on the word position information in the word based on the introduction of external lexical information to improve the recognition effect; Late-LSTM explicitly utilizes word and word sequence information. The gated recurrent unit allows the model to select the most relevant characters and words from the sentence for better NER results; Dict-BiLSTM-CRF merges dictionaries into a deep neural network to solve the medical text entity recognition task, proposing two different architectures and five different feature representation schemes to handle the task.

However, the various improved models mentioned above can learn the feature information of characters or words, but cannot handle the problem of multiple meanings of words, and the model performance improvement is limited. The BERT-BiLSTM-MHA-CRF model proposed in this paper, which can effectively solve the problem of multiple meanings of words and learn rich semantic feature information of characters, words and contexts, achieves excellent performance in the resume corpus and the CCKS2017 corpus with F1 values of 96.14% and 92.68%, respectively.

Gao M et al[25] proposed an attention-based ID-CNNs-CRF model that achieves excellent results in the CCKS2017 dataset. The paper further simplifies the text entity recognition process with the help of the Bert pre-training model, and also achieves excellent recognition results, providing a new way of thinking and direction for medical text processing.

4. **CONCLUSION.** For medical text entity recognition, the traditional language processing models cannot effectively represent the contextual semantic information in the text and cannot handle different contexts with multiple meanings of a word, which affects the effect of medical entity recognition. In the paper, we propose a full-word-based text entity recognition model ,just that a BERT-BiLSTM-MHA-CRF model, in which the BERT pre-processing language model generates word vectors that represent contextual semantic information, and the whole-word MASK processing mechanism is more suitable for processing Chinese text. The BERT preprocessing layer is able to generate word vectors characterizing contextual semantic information, the Bi-LSTM layer extracts bidirectional feature information from the input word vectors, the attention mechanism layer assigns weights to the output feature vectors to effectively obtain long-range dependencies in text utterances, and finally decodes them by CRF to generate entity label sequences. The experimental results show that the model achieves excellent performance on both the Resume corpus and the CCKS2017 corpus, with F1 reaching 96.14% and 92.68% respectively. The next step is to simplify the model structure and improve the model training

speed on the one hand, and to apply the model to other domains to complete the corresponding natural language processing tasks and improve the adaptability of the model on the other hand.

Acknowledgment. This work is partially supported by Natural Science Foundation of Anhui University of Chinese Medicine under Grant No.2020zrzd18, No.2019zrzd11,; Humanity Social Science foundation Grant 2021rwzd20, No.2020rwzd07; Anhui University of Chinese Medicine Quality Engineering Key Projects No. 2021zlgc046.

REFERENCES

[1] D. Zhang, W.L. Chen, Chinese named entity recognition based on contextually relevant word vectors, *Computer Science*, vol. 48, no. 3, pp. 233-238, 2021.
[2] Y. Nan, W. Pu, W. Zhang, Named Entity Recognition in Chinese Electronic Medical Records Based on Multi-Feature Integration, *Beijing Biomedical Engineering*, vol. 37, no. 3, pp. 279-284, 2018.
[3] Q.L. Li, J. Zheng, Text Sentiment Analysis Based on Hidden Markov Model, *Journal of Hangzhou Dianzi University (Natural Sciences)*, vol. 40, no. 6, pp. 50-55, 2020.
[4] L.Y. Lian, Named Entities Recognition of BiLSTM-CRF in Cyberspace Security Domain, *Journal of Heilongjiang University of Science and Technology*, vol. 30, no. 6, pp. 717-722, 2020.
[5] S.L. Zhu, Deep Learning Based Uyghur Named Entities Recognition, *Computer Engineering and Design*, vol. 40, no. 10, pp. 2874-2878, 2019.
[6] Q.T. Ren, H. Wang, X. Xiong, T. Fan, A Study on Construction and Application of Drama Term Extraction Model Integrating GCN Long-distance Constrain, *Data Analysis and Knowledge Discovery*, vol. 5, no. 6, pp. 1-19, 2021.
[7] H. Shen, Y.J. Zhang, B.H. Xie, Chinese Named Entity Recognition Based on BSTTC Model, *Computer Systems & Applications*, vol. 30, no. 6, pp. 262-270, 2021.
[8] Q. Yue, X. Li, Research on the Construction of Chinese Forestry Knowledge Graph Based on BERT and Bi-directional RNN, *Journal of Inner Mongolia University (Natural Science Edition)*, vol. 52, no. 2, pp. 176-184, 2021.
[9] T.Q. Jiang, L.X. Wang, Food Safety Named Entity Recognition Based on Bert and Attention Mechanism, *Science Technology and Engineering*, vol. 21, no. 3, pp. 1103-1108, 2021.
[10] P. Zhao, L.Y. Sun, Y. Wang, N. Ge, Chinese Scenic Spot Named Entity Recognition Based on BERT-BiLSTM-CRF, *Computer Systems & Applications*, vol. 29, no. 6, pp. 169-174, 2020.
[11] Y. Wang, M.X. Wang, S. Zhang, Y. Du, Alarm Text Named Entity Recognition Based on BERT, *Journal of Computer Applications*, vol. 40, no. 2, pp. 535-540, 2020.
[12] L.Y. Lian, S. Kong, J.W. Guo, Entity Recognition Method for Cyberspace Security Based on BERT, *Journal of Heilongjiang University of Science and Technology*, vol. 31, no. 3, pp. 394-398, 2021.
[13] K.K. Tseng, R. Zhang, C.M. Chen, M.M. Hassan, DNetUnet: a semi-supervised CNN of medical image segmentation for super-computing AI service, *The Journal of Supercomputing*, vol. 77, no. 4, pp. 3594-3615, 2021.
[14] A. Kumari, V. Kumar, M.Y. Abbasi, S. Kumari, P. Chaudhary, C.M. Chen, CSEF: Cloud-Based Secure and Efficient Framework for Smart Medical System Using ECC, *IEEE Access*, vol. 8, pp. 107838-107852, 2020.
[15] R. Farah, T. Shareef, Text Steganography based on Noorani and Darkness, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 12, no. 3, pp. 140-151, 2021
[16] X. Li, H. Yan, X. Qiu, X. Huang, FLAT: Chinese NER Using Flat-Lattice Transformer, *Proc of ACL*, pp. 6838-6842, 2020.
[17] Y. Chen, L. Fu, J. Zhang, S.L. Huang, Analysis method of unstructured resume information based on character sequence, *Computer Engineering and Design*, vol. 40, no. 6, pp. 1769-1774, 2019.
[18] R. Ma, M. Peng, Q. Zhang, Z. Wei, X. Huang, Simplify the usage of lexicon in Chinese NER, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5951-5960, 2020
[19] X.B. Ji, Y.H. Zhu, X. Xu, W.T. Liang, F. Zhan, Packaging named entity recognition based on attention mechanism, *Packaging Engineering*, vol. 40, no. 15, pp. 24-29, 2019.
[20] S.S. Zhong, X. Chen, M.H. Zhao, Y.J. Zhang, Incorporating Word-set Attention into Chinese Named Entity Recognition, *Journal of Jilin University (Engineering and Technology)*, 2021, https://doi.org/10.13229/j.cnki.jdxbgxb20200984.

[21] B. Ji, R. Liu, S.S. Li, J. Yu, Q.B. Wu, Y.S. Tan, J.J. Wu, A hybrid approach for named entity recognition in Chinese electronic medical record, *Bmc Medical Informatics and Decision Making*, vol. 19, no. 2, pp. 149-58, 2019.

[22] Q. Wang, Y. Xia, Y. Zhou, T. Ruan, D. Gao, P. He, Incorporating dictionaries into deep neural networks for the chinese clinical named entity recognition, arXiv preprint arXiv:1804.05017, 2018.

[23] Y. Zhang, X. Wang, Z. Hou, J. Li, Clinical named entity recognition from Chinese electronic health records via machine learning methods, *Jmir Medical Informatics*, vol. 6, no. 4, pp. e50, 2018.

[24] J. Hu, X. Shi, Z. Liu, X. Wang, Q. Chen, B. Tang, HITSZ_CNER: A hybrid system for entity recognition from Chinese clinical text, *Proc. CEUR Workshop*, vol. 1976, pp. 25-30, 2017

[25] M. Gao, Q. Xiao, S. Wu, K. Deng, An Improved Method for Named Entity Recognition and Its Application to CEMR, *Future Internet*, vol. 11, no. 9, pp. 185-199, 2019.