Research Article

Named Entity Recognition of Traditional Chinese Medicine Patents Based on BiLSTM-CRF

Na Deng [iD],[1] **Hao Fu** [iD],[1] **and Xu Chen** [iD][2]

[1]*School of Computer and Science, Hubei University of Technology, Wuhan (086)430068, China*
[2]*School of Information and Safety Engineering, Zhongnan University of Economics and Law, Wuhan (086)430073, China*

Correspondence should be addressed to Hao Fu; iamfuhao@foxmail.com

With the growing popularity of traditional Chinese medicine (TCM) in the world and the increasing awareness of intellectual property protection, the number of TCM patent application is growing year by year. TCM patents contain rich medical, legal, and economic information. Effective text mining of TCM patents is of great theoretical and practical significance (e.g., the R&D of new medicines, patent infringement litigation, and patent acquisition). Named entity recognition (NER) is a fundamental task in natural language processing and a crucial step before indepth analysis of TCM patent. In this paper, a method combining Bidirectional Long Short-Term Memory neural network with Conditional Random Field (BiLSTM-CRF) is proposed to automatically recognize entities of interest (i.e., herb names, disease names, symptoms, and therapeutic effects) from the abstract texts of TCM patents. By virtue of the capabilities of deep learning methods, the semantic information in the context can be learned without feature engineering. Experiments show that the BiLSTM-CRF-based method provides superior performance in comparison with various baseline methods.

1. Introduction

TCM has a long history and has been inherited for thousands of years. It is becoming increasingly popular all over the world for its mild medicinal properties and impressive therapeutic effects, especially for certain chronic and intractable diseases. On October 1, 2018, the World Health Organization brought TCM into the medical compendium with global influence, and remarkably, it has made significant contribution to the world's combat with the COVID-19 pandemic so far.

Patent is the carrier of the most advanced technology and method that contains rich technical, economic, and legal information. With the increasing awareness of intellectual property protection, the number of patent application is growing year by year. At present, patent texts, in parallel with scientific literature, have become an important source of knowledge in the era of the Internet, during which emerging technologies such as Web services, cloud computing, the Internet of Things, and wireless sensor networks have contributed to the protection, development, and promotion of

TCM [1–5]. It is expected to analyze patents as much as possible, so that the hidden valuable information can be extracted and utilized in the R&D of new medicines, patent infringement litigation, and patent acquisition. Since the R&D of TCM is a time-consuming and laborious process, if we can fully analyze the information in TCM patents, repeated medicine research will be largely avoided, R&D cycle will be shortened, and R&D costs will be saved. In this sense, the analysis of TCM patents is becoming a hot research topic. Before analyzing TCM patents, an essential step is to extract important named entities (e.g., herb names, disease names, symptoms, and therapeutic effects) from the TCM patent texts. These extracted entities can not only serve as the object of semantic extension in patent intelligent retrieval but also be the input in calculating patent text similarity.

Nowadays, the identification of these important named entities in TCM patent texts is faced with the following challenges:

(1) Due to the long history and wide geographical distribution of herbs, factors such as dialect and erroneous

records have led to a rather common phenomenon of aliases. One herb may have multiple names, and different herbs may share the same name. For example, "ganoderma lucidum" has synonyms such as "ChiZhi," "HongZhi," "MuLingZhi," "JunLingZhi," "WanNianXun," and "LingZhiCao," and both "black sesame" and "linseed" are known as "flax." Currently, there is not a complete Chinese herb thesaurus yet

(2) Due to different writing habits, TCM doctors and researchers tend to express a certain disease name in languages similar with classical Chinese, such as using "ShouHan" or "ShouLiang" to refer to "catch a cold"

(3) Due to the lack of labeled training sets of TCM patent texts, a proper training set has to be established manually while applying machine learning method for NER

At present, there are few studies on TCM patents mining, particularly in the NER of TCM patents. High-quality labeling of TCM patents still relies on manual work in a large extent. Therefore, we intend to design a model to fill the technical gap in NER of TCM patents. The main contributions of this paper are summarized as follows:

(1) We apply BiLSTM and CRF to the NER of TCM patents, which can comprehensively utilize the characteristics of the contexts

(2) The possible entity boundary problem in the task of NER is discussed, and practical solutions to the problem is proposed

The rest of the paper is organized as follows: Firstly, related work in the field of NER is presented, and then the models of CRF and BiLSTM are introduced, respectively. After that, the method proposed in this paper is introduced in detail, including the structure of the neural network and all key steps. Finally, the experiment and the analysis of experimental results are demonstrated.

2. Related Work

Named entity recognition (NER) is an important task in natural language processing that aims to identify key entities in the text. It paves way for applications such as information retrieval, relation extraction, machine translation, and question-answering system. Usually, the performance of these applications will depend directly on the accuracy of NER. To date, the research methods of NER basically fall into three categories, namely, dictionary-based, rule-based, and machine learning-based (deep learning included). Dictionary-based methods, widely used in the early stage, work by matching text against a dictionary (a list of entity names). This kind of method is simple, but the accuracy of recognition depends heavily on the accuracy of word segmentation, and it often fails when it comes to custom words. With regard to the rule-based method, extraction rules are summed up by experts after observing the characteristics of

the text, and entities are extracted using regular expressions. The downside is that the accuracy of this method depends on the quality of rules, and its portability is poor. At present, the most common method is based on machine learning or deep learning, which supports automatic characteristics learning of entities from the training set and has exhibited high generalization capability.

Lei et al. [6] investigated the effects of different types of features based on the CRF classifier and machine learning algorithms for NER in Chinese clinical text. The result showed that word segmentation based on a Chinese medical dictionary played a positive role.

Rahman et al. [7] designed a feature-oriented method that based on CRF to identify disease in biomedical literature. Despite its results needed to be improved further, sentence and token level features were given, thereby contributing to the overall performance and providing inspiration for our research.

In our previous research [8], the cotraining algorithm in weakly supervised machine learning was improved to recognize herb names in TCM patent texts. The method proposed does not require a large number of training samples and makes full use of the characteristics of the texts, including distribution and word-formation features. The deficiency of this method is that each iteration require manual participation, and high-quality candidate words have to be selected as the seeds for the next iteration. The recall rate also needs to be improved.

In recent years, deep learning has been developing rapidly since its emergence in the field of image recognition. Different from traditional machine learning methods, deep learning is characterized by automatic and effective feature extraction. Moreover, with the emergence of the Word2Vec method, deep learning-based methods have been applied to a variety of natural language processing tasks, such as NER, entity relationship extraction, part of speech analysis, and sentiment analysis.

Wei et al. [9] proposed an approach that combines CRF and Bi-RNN with SVM and outperformed the baseline by a large margin. The major factor that hindered the performance lies at redundant or missing modifier of the entity to be recognized. Likewise, the neural network architecture presented in Chen L's work [10] achieved good results on some corpora and outperformed the previously top performing systems; yet, boundary errors of word segmentation caused some performance loss. Ye et al. [11] examined the challenge of the lack of explicit labels in Chinese text to define the boundaries of words and proposed an NER model based on character-word vector fusion. The presented model was proved to reduce the dependence on the accuracy of the word segmentation algorithm and make the best use of the words' semantic characteristics. To address the problem of failing to correctly identify the boundary of an entity, we ensure the integrity and accuracy of the entity by bidirectional storage of text information and the introduction in of the BIO labeling method.

Based on experimental analysis, Wen et al. [12] pointed out that compared with traditional models, the deep learning-based methods can better capture the context

information and thus save the laborious work of feature selection. Furthermore, adding CRF layer over the neural network can bring better results. In a sentiment analysis experiment, Chen et al. [13] found that sentence classification can improve the performance of sentence-level sentiment analysis. These findings greatly benefit for our research, especially in the design of the proposed model structure.

As a special recurrent neural network (RNN) in deep learning, LSTM avoids the gradient disappearance problem of RNN [14] and is especially suitable for dealing with character sequences with dependency as it can efficiently utilize the characteristics of the context. Combining LSTM with CRF, which ensures the validity of the predicted label by applying rules it learns from the training dataset, we argue that promising recognition result is anticipated.

3. Conditional Random Field

Conditional random field, proposed by Lafferty et al. in 2001, is an undirected graph model and a statistical model that works well in labeling and segmenting structured data [15]. It is good at global normalization of sequence probability and can freely label for sequences by feature function, avoiding the conditional independence and solving the problem of labeling bias [16]. The conditional random field model randomly outputs variable Y under the condition of a given random variable X and constructs a conditional probability model $P(Y \mid X)$ that satisfies Markov property as shown in Equation (1):

$$P(Y_v \mid X, Y_w, w \neq v) = P(Y_v \mid X, Y_w, w \sim v). \quad (1)$$

In Equation (1), w represents all nodes except node v, $w \sim v$ represents all nodes connected with node v in an undirected graph, and Y_v and Y_w are the random variables corresponding to node v and w. The linear chain CRF [17] is frequently used in NER tasks. Given an observation sequence $X = \{X_1, X_2, \cdots, X_T\}$ and the corresponding label sequence $Y = \{Y_1, Y_2, \cdots, Y_T\}$, the conditional probability distribution $P(Y \mid X)$ of Y constitutes the conditional random field:

$$\begin{aligned} &P(Y_i \mid X, Y_1, Y_2, \cdots, Y_{i-1}, Y_{i+1}, \cdots, Y_T) \\ &= P(Y_i \mid X, Y_{i-1}, Y_{i+1}), i = 1, 2, \cdots, T. \end{aligned} \quad (2)$$

For a linear chain conditional random field, the loss function is shown in Equation (3) [18].

$$P(Y \mid X) = \frac{1}{Z(X)} \exp \left(\sum_{i=1}^{T} \sum_{k=1}^{K} w_k f_k(t, Y_t, Y_{t-1}, X) \right), \quad (3)$$

where $Z(X) = \sum_Y \exp \left(\sum_{i=1}^{T} \sum_{k=1}^{K} w_k f_k(t, Y_t, Y_{t-1}, X) \right). \quad (4)$

X represents the input sequence, Y_t represents the label of the current position, Y_{t-1} represents the label of the previ-

ous position, K represents the number of feature functions, T represents the size of X, $f_k(t, Y_t, Y_{t-1}, X)$ represents the kth feature function, w_k represents the weight of the kth feature function, and $Z(X)$ represents the normalization factor. Finally, the possibilities of all candidate labeling sequences are summed up. The conditional random field model uses the forward and backward algorithm to carry out the conditional probability and characteristic expectation of different sequence positions, uses the quasi-Newton method and other maximized likelihood estimations to solve model parameters, and uses the Viterbi algorithm to find out the optimal label sequence by dynamic programming.

4. Long Short-Term Memory Neural Network

Hochreiter and Schmidhuber [19] proposed long short-term memory (LSTM), which consists of memory units c_t, forget gates f_t, input gates i_t, and output gates o_t.

The structure detail of LSTM is shown in Figure 1. In LSTM, there are operations such as addition, multiplication, *tanh* function, and *sigmoid* function. A line from the input (x_t, h_{t-1}) to the output h_t represents the state of the cell. The output of this cell is a number between 0 and 1, where 0 means no passing at all and 1 means permitting to pass completely.

Firstly, what information to discard from the cell state will be decided. The forget gate f_t judges the importance of the past memories, that is, the extent to which the past memories participate in the generation of new memories, as shown in Equation (5).

$$f_t = \sigma \left(W_f[h_{t-1}, x_t] + b_f \right). \quad (5)$$

Next, what new information to store in the cell state will be determined. The input gate i_t determines how important the current word is, that is, to what extent it is helpful in generating new memories, as shown in Equation (6).

$$i_t = \sigma \left(W_i[h_{t-1}, x_t] + b_i \right). \quad (6)$$

At the same time, a tanh function is used to generate a new candidate memory unit $\tilde{c}_t$, as shown in Equation (7).

$$\tilde{c}_t = \tan h \left(W_f[h_{t-1}, x_t] + b_c \right). \quad (7)$$

Next, the cell state c_t is updated to obtain the memory unit of the current moment, as shown in Equation (8).

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t. \quad (8)$$

The final step is to determine the output of the model. The output gate o_t is first used to determine which part of the cell's state will be output, as shown in Equation (9).

$$o_t = \sigma \left(W_o[h_{t-1}, x_t] + b_o \right). \quad (9)$$

Then, the cell state at the current moment is processed through $\tan h$, and these two signals are considered
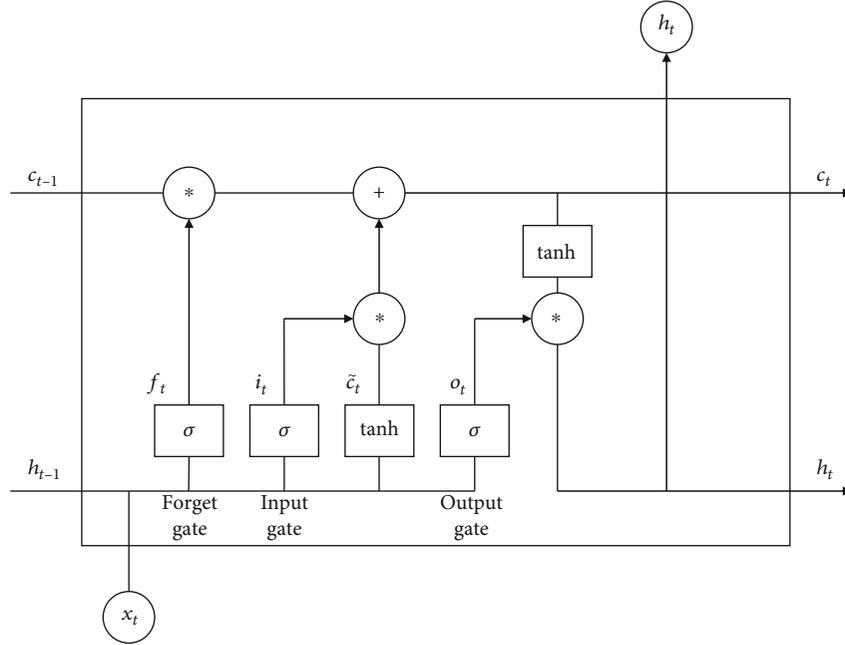
FIGURE 1: The structure of an LSTM cell [20].

comprehensively, so that only a certain part is output and the final output of the hidden layer is obtained.

$$h_t = o_t \times \tan h(c_t). \tag{10}$$

However, this kind of network only considers the influence of the past sequence at present, while ignoring the information later on and failing to achieve an ideal performance. Therefore, a bidirectional long short-term memory neural network (BiLSTM) model is introduced in, which can connect the output of the LSTM unit bidirectionally and capture bidirectional semantic dependencies, thereby improving the performance of the overall model.

5. The Proposed Model

In TCM patent texts, the description of a disease name usually has hints. For example, in the text sequence "DuiZhiLiaoLinBaJieYanJuYouXianZhuLiaoXiao (It has a significant effect on the treatment of lymphadenitis)", the word "ZhiLiao (treatment)" is often followed by a certain disease name. Therefore, when deciding whether a text contains a disease name or not, starting words play a key role in capturing the strong dependence in the context. On one hand, CNN tends to obtain static information of the text sequence "ZhiLiaoXiaoGuoMingXian (the therapeutic effect is notable)", and so when it receives the word "ZhiLiao (therapeutic)", it will mistakenly assume that the subsequent sequence "XiaoGuoMingXian (the effect is obvious)" is a disease name; on the other hand, although LSTM is good at obtaining long-term and long-range information in a sequence, its major weakness is that it cannot handle the noisy words in the rest of the sequence. For instance, given a text sequence "It can safely and effectively treat cough without toxic side effects and addiction", this sequence contains a disease name, but

there are noisy words both in the front part and the back part of the sequence. To sum up, CNN tends to capture keywords and thus cause misjudgment, and one-way LSTM is not sensitive enough to subsequent interference, leading to misjudgment as well. So, if only CNN or one-direction LSTM is used, interferences aforementioned may not be eliminated. However, BiLSTM can not only capture the information of the timing sequence dynamically but also make use of both the preceding and the subsequent information of the current word to ultimately obtain a strong dependency.

BiLSTM is well aware of the context information in the character sequence, and CRF helps improve the labeling accuracy at the sentence level. Combining the advantages of BiLSTM and CRF, the overall recognition accuracy will be improved. An illustrative structure of the BiLSTM-CRF model is shown in Figure 2.

The procedure of TCM entity recognition based on BiLSTM-CRF will be described in details later in this paper. Here, the core steps are listed as follows:

(1) Each character in TCM patent text will be mapped into a low-dimension dense vector by using a pretrained embedding matrix

(2) Embedding vector of each character will be taken as the input of each time step of BiLSTM Layer, and the hidden states of the forward and backward output will be spliced to obtain a complete hidden state sequence; that is, the context features of the text are extracted. The probability distribution matrix of the sequence and label is then calculated

(3) CRF Layer learns the potential relationship between sequences and excludes the label sequences that do not conform to the sentence-level grammatical rules. Finally, the optimal label sequence is ready for output
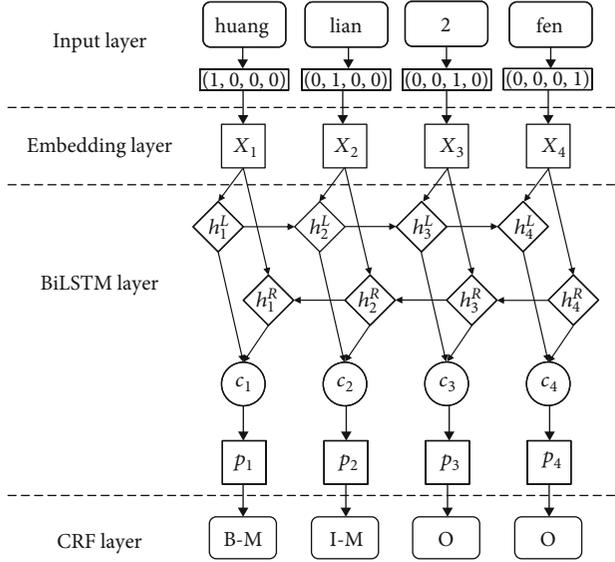
FIGURE 2: BiLSTM-CRF model structure based on Chinese characters.

5.1. Input Layer. At present, the input of most NER models is often words; so, the effect of entity recognition depends heavily on the effect of word segmentation. Characters, instead of words, contain a lot of linguistic information. In fact, studies have revealed that character-based NER generally performs better than word-based methods [21]. For this reason, this model takes characters as the initial input so that the accumulation of errors caused by poor segmentation can be mostly avoided. A global dictionary with size v is constructed, that is, a collection of v distinctive characters in the training set, in which each character corresponds to an identifier. A sentence containing n characters is one-hot encoded as an $n \times v$-dimension matrix, denoted as $W = (w_1, w_2, \cdots, w_n)$, where w_i represents the vector of the ith character of the sentence. A schematic diagram is shown in Figure 3.

5.2. Embedding Layer. TCM texts from the input layer are converted into vectors so that computers can calculate. The traditional one-hot representation cannot capture the semantic relationship between words, and it will also bring in dimensional disasters and data-sparse problem. Distributed representation methods, however, can map words into fixed-length, low-dimensional, dense vectors. The semantic similarity between words can be measured based on the distance of the words in the vector space, which well overcomes the shortcomings of one-hot representation [22]. Especially after Mikolov proposed the Word2Vec model in 2013, the distributed representation entered the practical stage, and the application of deep learning methods in the field of NLP reached a new height. Furthermore, in order to cope with the mismatch between the scale of the training dataset of the deep neural network model and the parameters to be trained, high-quality pretraining results were used to initialize the parameters to get better results. We fine-tune the initial word embeddings, modifying them during backpropagation

of the neural network model [23]. Each incoming identifier of a character w_i from the input layer is mapped into a d-dimension dense vector $X_i = (x_1, x_2, \cdots, x_d)$, where d is the embedding size that defines the number of features used to represent character. The vector is obtained according to Equation (11).

$$X_i = Hw_i, \tag{11}$$

where H denotes a $d \times v$-dimension pretrained weight matrix. The vector is then transmitted to BiLSTM Layer as the input. At this point, a sentence containing n characters is mapped into a dense matrix X (with the shape of $n \times d$) from the initial sparse one-hot matrix W (with the shape of $n \times v$).

5.3. BiLSTM Layer. This layer is used to extract sentence features. As shown in Figure 1, every character vector $X_i = (x_1, x_2, \cdots, x_d)$ is taken as the input of BiLSTM Layer, and the output hidden state of the ith character $h_i^L = (h^l_1, h_2^l, \cdots, h_d^l)$ on the left is obtained through the forward LSTM. Likewise, after backward LSTM, the output hidden state $h_i^R = (h^r_1, h_2^r, \cdots, h_d^r)$ on the right is obtained. h_i^L and h_i^R are then spliced to obtain the output hidden state $c_i = [(h_i^L, h_i^L)]$ of each position i, and finally, the sentence-level hidden state sequence $C = (c_1, c_2, \cdots, c_n)$ is obtained. A full connection layer is used to map the hidden state vector to k dimensions, where k is the number of labels in the label set, so as to extract features and provide a probability distribution matrix $P = (p_1, p_2, \cdots, p_n)$. Element $p_{i,j}$ represents the probability of classifying the character w_i as the jth label. The pseudocode of the BiLSTM Layer is shown in Algorithm 1.

5.4. CRF Layer. This layer carries out sentence-level sequence labeling to ensure the generation of the globally optimal labeling sequence. The output of the BiLSTM Layer p_i is independent of each other, ignoring the strong dependence between its preceding label p_{i-1} and its subsequent label p_{i+1}. The CRF layer can automatically obtain some restrictive rules from the training data and conduct sentence-level adjustment, which will reduce the probability of illegal sequences and improve the accuracy of label sequence prediction. An intuitive explanation will be given later in this section in conjunction with the mathematical notation. Considering the need to add a start state at the beginning of a sentence and a stop state at the end, the data structure of the CRF layer is a $(k+2) \times (k+2)$ state transfer matrix A. $A_{i,j}$ represents the transfer score from the ith label to the jth label. For a sentence $W = (w_1, w_2, \cdots, w_n)$ as the input, assuming that a predicted label sequence $y = (y_1, y_2, \cdots, y_n)$ is obtained, the score of the prediction is defined as

$$s(W, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i}, \tag{12}$$

where P_{i, y_i} is the probability of the ith position of BiLSTM outputs being y_i, and $A_{y_i, y_{i+1}}$ is the transfer probability from
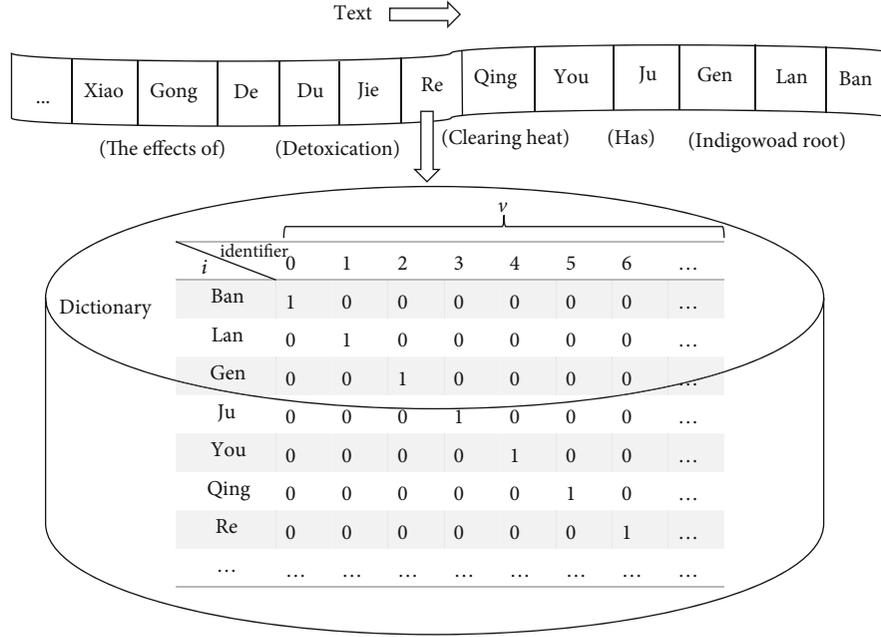
FIGURE 3: Character-based text input and the construction of the global dictionary.

Input: Pre-trained character embedding X.
Output: Probability distribution matrix P of the input sequence
(1)   Step 1: The character vectors from X are sent into the forward LSTM layer
(2)   for $i \in$ length (X) do
(3)   send X_i to BiLSTM Layer
(4)   end for
(5)   Step 2: The state of the cell in the current LSTM network is updated
(6) $f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)$
(7) $i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)$
(8) $\widetilde{c_t} = \tan h(W_f[h_{t-1}, x_t] + b_c)$
(9) $c_t = f_t * c_{t-1} + i_t * \widetilde{c_t}$
(10) $o_t = \sigma(W_o[h_{t-1}, x_t] + b_o)$
(11) $h_t = o_t * \tan h(c_t)$
(12)   Step 3: The character vectors from X are sent into the backward LSTM layer and the above 2 steps are repeated
(13)   Step 4: The forward and backward sequences of hidden layers are spliced to obtain a sentence-level hidden state sequence C rich in contextual information
(14)   Step 5: C is sent into a full connection layer and the prediction matrix P is obtained
(15)   Return P;

ALGORITHM 1

y_i to y_{i+1}. The score of a candidate sequence is jointly determined by the features P extracted from the BiLSTM layer and the transfer matrix A aforementioned. Suppose, for instance, a sequence that the BiLSTM layer outputs is "B-M, B-M, I-M, O," but the probability of "B-M, B-M" in the transfer matrix A is quite small or even negative (in a practical sense, the chances of it happening are also slim to none); so, the score s will be reduced, and consequently, this unreasonable prediction sequence is likely to be ruled out. For each training sample W, the Viterbi algorithm [24] is used to calculate the score $s(W, y)$ of all possible labeled sequence y, and then a *softmax* function is added to normalize all the scores. Equation (13) is used to calculate the probability of sequence y.

$$p(y \mid W) = \frac{e^{s(W,y)}}{\sum_{y \in Y_W} e^{s(W,y)}}. \tag{13}$$

During model training, for the sentence input sequence X, the loss function is set to take the logarithm of the probability of the target true label sequence Y. To maximize the probability corresponding to the true label sequence, the method of taking the negative value and then minimizing
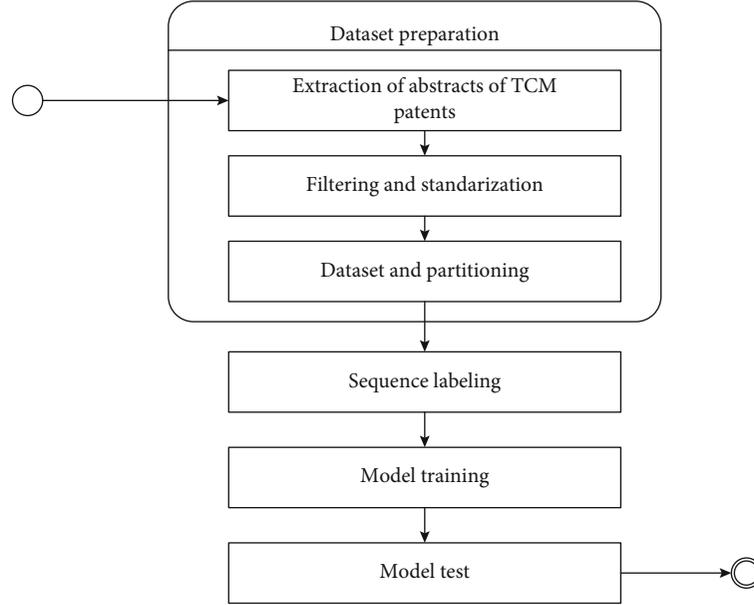
FIGURE 4: Named entity recognition process of TCM patent abstracts.

TABLE 1: Examples of each entity category.

Entity category	Examples of sample
Herb name	GanCao (liquorice), FuLing (Poria cocos), LianQiao (Forsythia), HuangQi (Astragalus membranaceus), DangGui (Angelica sinensis), BaiZhu (Rhizoma Atractylodis)
Disease name	GaoNiaoSuanXueZheng (hyperuricemia), GuoMinXingBiYan (anaphylactic rhinitis), TangNiaoBing (diabetes), YaZhouYan (periodontitis), WeiKuiYang (gastric ulcer), ZhiChuang (hemorrhoids)
Symptom	HuangHeBan (chloasma), FuXie (diarrhea), ShiYuBuZhen (loss of appetite), KouKe (thirst), ChuXue (bleeding), PiLao (fatigue)
Therapeutic effect	RunFei (nourishing lung), JianPi (tonifying spleen), JiangZhi (lipid-lowering), AnShen (soothe the nerves), ZhiKe (relieving a cough), XiaoZhong (detumescence)

it is adopted, and the gradient descent algorithm is introduced in to solve the parameters and maximize the log-likelihood function:

$$\log\left(p(Y \mid X)\right) = s(X, Y) - \log\left(\sum_{Y \in Y_x} e^{s(X,Y)}\right). \quad (14)$$

In the process of prediction, scores s corresponding to every candidate sequence y are calculated according to the trained parameters, and the Viterbi algorithm using dynamic programming at its core is used to calculate the optimal path. The predicted result is denoted as Y^*:

$$Y^* = \mathrm{argmax}_{Y \in Y_X}(s(X, Y)). \quad (15)$$

In summary, a BiLSTM-CRF model was established in this paper to identify the entities of herb names, disease names, symptoms, and therapeutic effects in TCM patent texts. Given a text sequence as input, the model can output relevant entities in it. Combined with word vector technology, LSTM has obvious superiority when dealing with text sequence as it can take full advantage of long-term and

TABLE 2: BIO labeling set for CRF.

Label of entity	Beginning label	Intermediate or end label
Herb name	B-M	l-M
Disease name	B-D	l-D
Symptom	B-S	l-S
Therapeutic effect	B-E	l-E
Nonentity	O	O

long-distance information dependencies [25]. BiLSTM can provide even more comprehensive contextual information and makes it easier to learn about contextual dependencies. The additional CRF layer makes up for the deficiency of the BiLSTM model by optimizing the recognition results comprehensively from the sentence level.

6. Experiment

6.1. Experimental Procedure. Based on the BiLSTM-CRF algorithm, this paper has completed the NER experiment on TCM patent texts. The main experimental process is shown in Figure 4.

TABLE 3: Instances of sequence labeling.

No	State	Form
1	1	JinYinHua3-5Fen, AiYe2-5Fen, YuXingCao2-5Fen, HuangLian2-4Fen (3-5 portions of honeysuckle, 2-5 portions of mugwort, 2-5 portions of houttuynia cordata, and 2-4 portions of coptis)
	2	JinYinHua/M 3-5Fen, AiYe/M 2-5Fen, YuXingCao/M 2-5Fen, HuangLian/M 2-4Fen
	3	Jin[B-M]yin[I-M]Hua[I-M]3[O]-[O]5[O]fen[O],[O]Ai[B-M]Ye[I-M]2[O]-[O]5[O]fen[O],[O]Yu[B-M]Xing[I-M]Cao[I-M]2[O]-[O]5[O]fen[O],[O]Huang[B-M]Lian[I-M]2[O]-[O]4[O]fen[O]
2	1	DuiManXingLeiFengShiDengYinQiDeGanWeiBuHe, XiaoHuaBuLiang, ShiYuBuZhen, XiongFuZhangManYouLiangHaoXiaoGuo (it has good effect on liver and stomach disharmony caused by chronic rheumatism, indigestion, inappetence, chest, and abdomen distension)
	2	DuiManXingLeiFengShi/D DengYinQiDeGanWeiBuHe/S, XiaoHuaBuLiang/S, ShiYuBuZhen/S, XiongFuZhangMan/S YouLiangHaoXiaoGuo
	3	Dui[O]man[B-D]Xing[I-D]Lei[I-D]Feng[I-D]Shi[I-D]Deng[O]yin[O]qi[O]De[O]Gan[B-S]Wei[I-S]Bu[I-S]He[I-S],[O]Xiao[B-S]Hua[I-S]Bu[I-S]Liang[I-S],[O]Shi[B-S]Yu[I-S]Bu[I-S]Zhen[I-S],[O]Xiong[B-S]Fu[I-S]Zhang[I-S]man[I-S]you[O]Liang[O]Hao[O]Xiao[O]Guo[O]
3	1	YiZhongZhiLiaoZhiChuangDeZhongYaoZuHeWu, GaiZuHeWuBaoKuoXuChangQing, DiHuang, WuBeiZi, ShengMa, HuaiShanYao, KuShen, HuangQi (the invention relates to a TCM composition for the treatment of hemorrhoids, which includes Xu Changqing, rehmannia glutinosa, gallnut, cohosh, Huai yam, sophora flavescens, and astragalus)
	2	YiZhongZhiLiaoZhiChuang/D DeZhongYaoZuHeWu, GaiZuHeWuBaoKuoXuChangQing/M, DiHuang/M, WuBeiZi/M, ShengMa/M, HuaiShanYao/M, KuShen/M, HuangQi/M
	3	Yi[O]Zhong[O]Zhi[O]Liao[O]Zhi[B-D]Chuang[I-D]De[O]Zhong[O]Yao[O]Zu[O]He[O]Wu[O],[O]Gai[O]Zu[O]He[O]Wu[O]Bao[O]Kuo[O]Xu[B-M]Chang[I-M]Qing[I-M],[O]Di[B-M]Huang[I-M],[O]Wu[B-M]Bei[I-M]Zi[I-M],[O]sheng[B-M]ma[I-M],[O]Huai[B-M]Shan[I-M]Yao[I-M],[O]Ku[B-M]Shen[I-M],[O]Huang[B-M]qi[I-M]
4	1	BenFaMingYaoFangJuYouQingReXieHuo, XiaoZhongZhiTongZhiGongXiao, LinChuangShangYongYuZhiLiaoYaYinYan, YaZhouYan, YaSuiYan, GuanZhouYanJiQiSuoYinQiDeYaTongDengJiBing (the prescription of the invention has the effect of clearing away heat, reducing swelling, and relieving pain and is clinically used to treat gingivitis, periodontitis, pulpitis, pericoronitis, and other diseases caused by toothache.)
	2	BenFaMingYaoFangJuYouQingRe/E XieHuo/E, XiaoZhong/E ZhiTong/E ZhiGongXiao, LinChuangShangYongYuZhiLiaoYaYinYan/D, YaZhouYan/D, YaSuiYan/D, GuanZhouYan/D JiQiSuoYinQiDeYaTong/D DengJiBing
	3	Ben[O]Fa[O]Ming[O]Yao[O]fang[O]Ju[O]you[O]Qing[B-E]re[I-E]Xie[B-E]Huo[I-E],[O]Xiao[B-E]Zhong[I-E]Zhi[B-E]Tong[I-E]Zhi[O]gong[O]Xiao[O],[O]Lin[O]Chuang[O]Shang[O]Yong[O]Yu[O]Zhi[O]Liao[O]Ya[B-D]yin[I-D]Yan[I-D],[O]Ya[B-D]Zhou[I-D]Yan[I-D],[O]Ya[B-D]sui[I-D]Yan[I-D],[O]guan[B-D]Zhou[I-D]Yan[I-D]Ji[O]qi[O]Suo[O]yin[O]qi[O]De[O]Ya[B-D]Tong[I-D]Deng[O]Ji[O]Bing[O]
5	1	BenFangSuoYongYaoHuoXue, GuJing, XieHuoDeLongGu, ShiGaoWeiJunYao; ZiBu, GuTuo, ShengJin, YangYanDeRenShen, FuLing, MeiGuiWeiChenYao; YiQuFeng, KangJunZhiYangDePuGongYing, FangFeng, BoHeWeiZuoYao; BaiZhu, SanQi, GanCaoWeiShiYao (the medicine used in this prescription includes keel and gypsum for invigorating blood and strengthening essence: ginseng, poria and rose for nourishing, solidifying, and nourishing body fluid; dandelion and peppermint for dispelling wind, killing bacteria, and stopping itching. Baizhu, panax notoginseng, and licorice are also included.)
	2	BenFangSuoYongYaoHuoXue/E, GuJing/E, XieHuo/E DeLongGu/M, ShiGao/M WeiJunYao; ZiBu/E, GuTuo/E, ShengJin/E, YangYan/E DeRenShen/M, FuLing/M, MeiGui/M WeiChenYao; YiQuFeng/E, KangJun/E ZhiYang/E DePuGongYing/M, FangFeng/M, BoHe/M WeiZuoYao; BaiZhu/M, SanQi/M, GanCao/M WeiShiYao
	3	Ben[O]fang[O]Suo[O]Yong[O]Yao[O]Huo[B-E]Xue[I-E],[O]Gu[B-E]Jing[I-E],[O]Xie[B-E]Huo[I-E]De[O]long[B-M]Gu[I-M],[O]Shi[B-M]Gao[I-M]Wei[O]Jun[O]Yao[O]; [O]Zi[B-E]Bu[I-E],[O]Gu[B-E]Tuo[I-E],[O]sheng[B-E]Jin[I-E],[O]Yang[B-E]Yan[I-E]De[O]Ren[B-M]Shen[I-M],[O]Fu[B-M]Ling[I-M],[O]Mei[B-M]Gui[I-M]Wei[O]Chen[O]Yao[O]; [O]Yi[O]Qu[B-E]Feng[I-E],[O]Kang[B-E]Jun[I-E]Zhi[B-E]Yang[I-E]De[O]Pu[B-M]gong[I-M]Ying[I-M],[O]fang[B-M]Feng[I-M],[O]Bo[B-M]He[I-M]Wei[O]Zuo[O]Yao[O]; [O]Bai[B-M]Zhu[I-M],[O]san[B-M]qi[I-M],[O]Gan[B-M]Cao[I-M]Wei[O]Shi[O]Yao[O]

6.2. Dataset Preparation. Since there is currently no public dataset for experiments in the field of TCM patents, we first crawled the TCM patents' abstract texts from intellectual property websites, a total of 2000 copies.

The data collected by crawlers contained a large number of nontext structure data such as website tags, links, and special characters, which was not conducive to sequence labeling [26]. Through technologies such as regular expressions and character format standardization, nontext data was removed, and 1600 abstract texts were finally preserved as experimental corpus.

The dataset was then partitioned into three parts, i.e., training set, verification set, and test set. The proportion of these three parts was 6 : 2 : 2 and specifically, 147,788 Chinese

Table 4: Named entity recognition model parameter settings.

Parameter	Value
Hidden size	128
Learning rate	0.001
Batch size	64
Max epoch	50

characters were used as the training set, 48,889 as the verification set and 49,293 as the test set.

6.3. Sequence Labeling. In simple terms, sequence labeling is to mark each element in the given sequence with a corresponding label so that it is possible for the neural network to learn. According to the selection criteria, 1600 high-quality abstract texts were selected for this experiment. The model designed in this paper is mainly aimed at identifying four categories of entities, namely, herb names, disease names, symptoms, and therapeutic effects in TCM patent texts. Examples of each entity class are shown in Table 1.

Firstly, each entity of the four categories should be identified. Secondly, the BIO (Begin, Intermediate, Other) labeling method [27] is adopted for text labeling, where "B" means the first character of a word, "I" means the nonfirst character of the word, and "O" means a nonfocus character or punctuation. In the specific labeling process, the labeling methods shown in Table 2 are used to distinguish between entities, where the initial character of a herb name is expressed in the form of B-M (Begin-herb), and the rest of the word is expressed in the form of I-M (Intermediate or end-herb).

Table 3 shows the detailed process of TCM patent sequence labeling. Take sample No.1 for example, the original text is denoted as state 1 in the table. With the help of a corpus labeling tool YDEEA [28], we need to find the entities in the text and then label them with corresponding marks. For example, as shown in state 2, "JinYinHua/M" represents that "JinYinHua" is a herb name entity. The tool will automatically generate a labeled version of the text for us in the form shown in state 3.

Finally, the labeled characters are directly sent to the input layer for processing; that is, Word2Vec embedding method is used to generate a 128-dimension word vector matrix for model training.

6.4. Model Training. PyTorch is an open-source Python machine learning library developed by Facebook Artificial Intelligence Institute (FAIR). It is widely used to implement various machine learning algorithms. The BiLSTM-CRF model proposed in this paper was implemented in the PyTorch framework. Based on the training experience of relevant papers [29] and through multiple adjustments, the main parameters considered in the experiment are finally set as shown in Table 4.

6.5. Model Test. We evaluate experimental results by comparing the recognition results of different models with manual labeling as bench mark. Precision rate P, recall rate R, and

$F1$ value are selected as evaluation indicators [30], as shown in Equation (16) to (18).

$$P = \frac{correct_{num}}{actual_{num}}, \qquad (16)$$

$$R = \frac{correct_{num}}{true_{num}}, \qquad (17)$$

$$F1 = 2 \times P \times R/(P + R). \qquad (18)$$

$actual_{num}$, $true_{num}$, and $correct_{num}$ represent the actual number of entities identified in the sample, the number of true entities, and the number of entities correctly identified, respectively. An example is given in Figure 5 to illustrate the calculation of the three indicators in discussion.

In this simple case, $true_{num}$ is 2 (a herb name and a therapeutic effect), $actual_{num}$ is 1 (a herb name), and $correct_{num}$ is also 1 (a herb name). Substitute these values into Equation (16) to (18) and we can conclude that the values of P, R, and $F1$ are equal to 100.00%, 50.00%, and 66.67%, respectively. According to the definition, The precision rate represents how many entities among the prediction result are true entities, while the recall rate represents how many entities among the original samples are correctly predicted. Since these two indicators reflect the recognition performance from different perspectives, their weighted harmonic mean needs to be considered for a comprehensive evaluation. The most commonly used is the F1 value, which indicates that the method is effective when the F1 value is high.

7. Results and Discussion

The experimental results are the average scores of the optimal model, as shown in Table 5: precision rate is 94.63%, recall rate is 94.47%, and F1 value is 94.48%. Based on the analysis of experimental results and labeled corpus, it can be seen that the performance of entity recognition, in a large part, is determined by the number of labels and to what extent different categories of entities are distinguishable from one another.

The category of herb names is labeled in the largest amount, and it is also unlikely to be recognized as any other category; so, it makes sense that its recognition result outperforms other categories. Although there are fewer training samples for therapeutic effects and disease names, their degrees of distinction are relatively high; so, their recognition result turns out favorable as well. Symptom entities are the least labeled and mostly likely to be confused with entities in other categories. For example, "ZhuiJianPanTuChu (intervertebral disc herniation)" is possibly regarded as a symptom or as a disease name, which seriously reduces the accuracy of identification. Furthermore, the boundary of symptoms is not clearly defined, and its common to have a symptom nested within a therapeutic effect. For example, if "JingMaiXueShuan (venous thrombosis)" in "KangJingMaiXueShuan-XingCheng (to alleviate venous thrombosis)" is labeled as a symptom entity while the model recognizes the entire phrase as a therapeutic effect, a false recognition will occur. Thus,
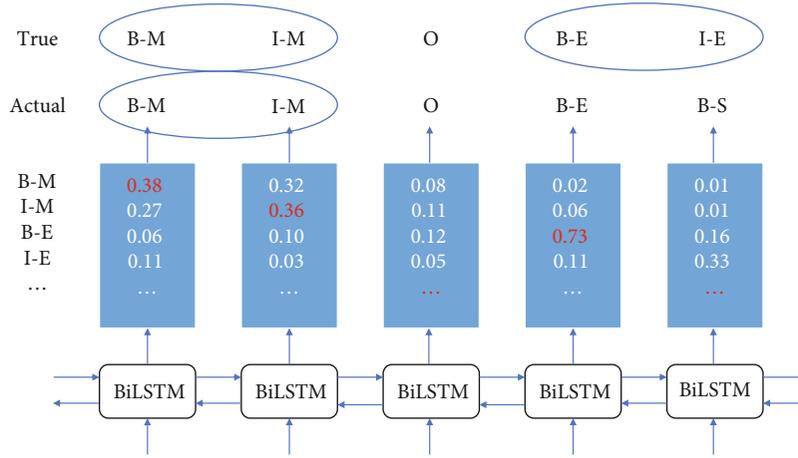
FIGURE 5: Experimental evaluation based on the output of the model.

TABLE 5: Experimental results.

Entity	Precision (%)	Recall (%)	F1 value (%)
Herb name	95.65	93.31	94.47
Disease name	78.05	69.11	70.45
Symptom	69.46	40.99	51.56
Therapeutic effect	78.94	78.94	78.94
Nonentity	97.62	97.28	97.45
Average	94.63	94.47	94.48

TABLE 6: Comparisons between related models.

Model	Precision (%)	Recall (%)	F1 value (%)
HMM	93.20	92.43	92.69
LSTM	93.29	93.13	93.09
BiLSTM	93.65	93.86	93.66
BiLSTM-CRF	94.63	94.47	94.48

phenomenon of nesting will result in poor accuracy of symptoms' recognition.

7.1. Model Performance Comparison. To verify the effective of the proposed model, we conducted comparative experiments on several relevant baseline models under the same environment, and the results are presented in Table 6. It can be seen that the recognition results of BiLSTM and BiLSTM-CRF models exceeded HMM in accuracy, recall rate, and F1 value, indicating that the performance of the BiLSTM-CRF model in this experiment is superior to that of the traditional machine learning algorithm HMM. This result can be explained from at least two perspectives. On one hand, in terms of word' representation, HMM involves one-hot representation while RNN uses a kind of distributed representation; so, the latter is more efficient in the face of high dimensions; On the other hand, from the perspective of the evolution mode of the hidden state, BiLSTM, as an RNN based model, replaces the linearity of HMM with a highly nonlinear one; so, its expression ability is stronger.

When the sample size goes up, the recognition effect of the LSTM method shows an impressive boost, thanks to its full consideration of timing characteristics [31]. Compared with the HMM model, the LSTM model uses a deeper and more complex neural network. On the basis of the ordinary neural network structure, the recurrent neural network structure is integrated, and timing characteristics are further considered. Therefore, this model is suitable for dealing with the contextual relationship in texts where the output at each time step is affected by the states of the previous time step.

In addition, from LSTM to BiLSTM, the overall F1 value increased from 93.09% to 93.66%, which is an apparent improvement. This is because LSTM only extracts one-way features of sequences, resulting in the lack of many useful features that make sense for sequence labels, whereas BiLSTM can extracts the features from both forward and backward directions of the sequence, so as to obtain the knowledge more comprehensively and thereby achieve better performance [32].

Moreover, by adding the CRF layer, the BiLSTM-CRF model obtained the optimal probability of label transfer. Compared with BiLSTM, precision, recall, and F1 value of the proposed model were improved by 0.98%, 0.61%, and 0.82%, respectively.

From the comprehensive recognition results, there is still room for further improvement of the method at present. The performance of the proposed model is mainly restricted by two factors:

(1) The scale of training data. Current data scale cannot well support the parameters required by the model, resulting in a negative impact on the learning of the model

(2) Entity labeling granularity. As discussed in the previous section, the entity categories used in this paper lead to noticeable nesting between symptoms and therapeutic effects. For example, the symptom "Ke (cough)" is nested in the therapeutic effect entity "ZhiKe (to relieve a cough)". In fact, in this case, we can simply label "cough" as a symptom and words such as "stop", "eliminate," and "relieve" only serve as clue words indicative of a symptom entity so that

the possibility of intermixing between entities can be effectively reduced

In the future, with the development of transfer learning, unsupervised learning technology, and radical vector feature representation methods that are more granular than word vectors, better recognition results are expected to be achieved with smaller data sizes and fewer labels [33]. Furthermore, the combination of coarse and fine-grained methods, together with entity alignment and semantic disambiguation technology, may address the problem of polysemy in TCM texts and improve the recognition performance.

8. Conclusions

In this paper, a BiLSTM-CRF model is constructed to obtain the bidirectional semantic features of the context and identify four types of entities, namely, herb names, disease names, symptoms, and therapeutic effects. Favorable results have been achieved, with the overall F1 value reaching 94.48%. The significant advance in comparison with HMM, LSTM, and BiLSTM indicates that the model constructed in this paper on NER is able to provide strong support for subsequent natural language processing applications and provide theoretical and technical reference for researchers in relevant fields. Besides, in view of the strong portability of the deep learning model, it can be applied to the NER tasks in various fields other than TCM patents as in our research. For any available standard dataset in a certain field, we can define a set of entity types properly, train the model, and extract the target entities for further analysis and utilization. Yet, there is still room for improvement in the task of NER in TCM texts. In future research, the following aspects shall be taken into account. On one hand, a corpus with higher purity, richer content, and larger scale is recommended to build higher quality character vectors and closer attention to the standardization of dataset labeling that is needed; On the other hand, for the rather ubiquitous phenomenon of nesting among entity categories in TCM texts, the granularity of entity labeling can be further optimized by referring to the solution presented in the previous section.

Data Availability

The dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

References

[1] A. Sangaiah, M. Suraki, M. Sadeghilalimi, S. Bozorgi, A. Hosseinabadi, and J. Wang, "A new meta-heuristic algorithm for solving the flexible dynamic job-shop problem with parallel machines," *Symmetry*, vol. 11, no. 2, p. 165, 2019.

[2] A. K. Sangaiah, A. A. R. Hosseinabadi, M. B. Shareh, S. Y. B. Rad, A. Zolfagharian, and N. Chilamkurti, "Iot resource allocation and optimization based on heuristic algorithm," *Sensors*, vol. 20, no. 2, p. 539, 2020.

[3] A. K. Sangaiah, M. Sadeghilalimi, A. A. R. Hosseinabadi, and W. Zhang, "Energy consumption in point-coverage wireless sensor networks via bat algorithm," *IEEE Access*, vol. 7, pp. 180258–180269, 2019.

[4] A. K. Sangaiah, G.-B. Bian, S. M. Bozorgi, M. Y. Suraki, A. A. R. Hosseinabadi, and M. B. Shareh, "A novel quality-of-service-aware web services composition using biogeography-based optimization algorithm," *Soft Computing*, vol. 24, pp. 8125–8137, 2019.

[5] S. M. G. Kashikolaei, A. A. R. Hosseinabadi, B. Saemi, M. B. Shareh, A. K. Sangaiah, and G.-B. Bian, "An enhancement of task scheduling in cloud computing based on imperialist competitive algorithm and firefly algorithm," *The Journal of Supercomputing*, vol. 76, pp. 6302–6329, 2019.

[6] J. Lei, B. Tang, X. Lu, K. Gao, M. Jiang, and H. Xu, "A comprehensive study of named entity recognition in Chinese clinical text," *Journal of the American Medical Informatics Association*, vol. 21, no. 5, pp. 808–814, 2014.

[7] H. Rahman, T. Hahn, and R. Segall, "Advanced feature-driven disease named entity recognition using conditional random fields," in *the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pp. 469–469, New York, NY, USA, 2016.

[8] N. Deng and C. Xiong, "Serialized co-training-based recognition of medicine names for patent mining and retrieval," *International Journal of Data Warehousing and Mining*, vol. 16, no. 3, pp. 87–107, 2020.

[9] Q. Wei, T. Chen, R. Xu, Y. He, and L. Gui, "Disease named entity recognition by combining conditional random fields and bidirectional recurrent neural networks," *Database*, vol. 2016, article baw140, 2016.

[10] C. Lyu, B. Chen, Y. Ren, and D. Ji, "Long short-term memory RNN for biomedical named entity recognition," *BMC Bioinformatics*, vol. 18, no. 1, p. 462, 2017.

[11] N. Ye, X. Qin, L. Dong, X. Zhang, and K. Sun, "Chinese named entity recognition based on character-word vector fusion," *Wireless Communications and Mobile Computing*, vol. 2020, Article ID 8866540, 7 pages, 2020.

[12] Y. Wen, C. Fan, G. Chen, X. Chen, and M. Chen, "A survey on named entity recognition," *Communications, Signal Processing, and Systems*, vol. 571, pp. 1803–1810, 2020.

[13] T. Chen, R. Xu, Y. He, and X. Wang, "Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN," *Expert Systems with Applications*, vol. 72, pp. 221–230, 2017.

[14] T. Mikolov, S. Kombrink, L. Burget, J. Cernocky, and S. Khudanpur, "Extensions of recurrent neural network language model," in *IEEE International Conference on Acoustics,*

Speech and Signal Processing, pp. 5528–5531, Prague, Czech Republic, 2011.

[15] J. Lafferty, A. Mccallum, and F. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *Proceedings of the 18th International Conference on Machine Learning*, pp. 282–289, Williamstown, MA, USA, 2001.

[16] K. Zhang, X. Wang, J. Ren, and C. Liu, "Efficiency improvement of function point-based software size estimation with deep learning model," *IEEE Access*, 2020.

[17] C. Sutton and A. Mccallum, *An Introduction to Conditional Random Fields for Relational Learning*, Now Publishers Inc., 2007.

[18] L. Wu, L. Liu, H. Li, and Y. Gao, "A Chinese toponym recognition method based on conditional random field," *Journal of Wuhan University (information science edition)*, vol. 42, no. 2, pp. 150–156, 2017.

[19] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[20] Y. Gu, J. Huo, H. Yang, Y. Lu, and Y. Guo, "Chinese named entity recognition method in motor field based on BERT," *Computer Engineering*, pp. 1–7, 2020.

[21] P. Yang and W. Dong, "Chinese named entity recognition method based on BERT embedding," *Computer Engineering*, vol. 46, no. 4, pp. 40–45+52, 2020.

[22] X. Gao, C. Du, and J. Wu, "Research on method of military named entity recognition based on BiLSTM-CRF," *Radio Engineering*, vol. 50, no. 12, pp. 1050–1054, 2020.

[23] T. Wang, P. Xuan, Z. Liu, and T. Zhang, "Assistant diagnosis with Chinese electronic medical records based on CNN and BiLSTM with phrase-level and word-level attentions," *BMC Bioinformatics*, vol. 21, no. 1, p. 230, 2020.

[24] S. Benedetto and E. Biglieri, "Viterbi algorithm," in *Principles of Digital Transmission*, Springer, 2002.

[25] J. Hu and X. Zheng, "Opinion extraction of government microblog comments via BiLSTM-CRF model," in *2020 ACM/IEEE Joint Conference on Digital Libraries*, pp. 473–475, Wuhan, China, 2020.

[26] P. Zhao, C. Zhao, H. Wu, and W. Wang, "Research on named entity recognition of Chinese agricultural based on attention mechanism," *Journal of Agricultural Machinery*, pp. 1–13, 2021.

[27] N. N. Xue and S. P. Converse, "Combining classifiers for Chinese word segmentation," in *First Sighan Workshop Attached with Coling*, pp. 1–7, Taipei, China, 2002.

[28] A. Maimai, S. Wushou, M. Palidan, and W. Yang, "Uyghur named entity recognition based on BiLSTM CNN-CRF model," *Computer Engineering*, vol. 44, no. 8, pp. 230–236, 2018.

[29] Q. Zeng, W. Xiong, J. Du, N. Bin, and R. Guo, "Electronic medical record named entity recognition combined with self-attention BiLSTM-CRF," *Computer Applications and Software*, vol. 38, no. 3, pp. 159–162, 2021.

[30] P. Tang, P. Yang, Y. Shi, Y. Zhou, F. Lin, and Y. Wang, "Recognizing Chinese judicial named entity using BiLSTM-CRF," *Journal of Physics: Conference Series*, vol. 1592, no. 1, article 12040, 2020.

[31] Z. Zhang, H. Liu, L. Ding, B. Wu, and J. Yu, "Identifying moves of research abstracts with deep learning methods," *Data Analysis and Knowledge Discovery*, vol. 3, no. 12, pp. 1–9, 2019.

[32] W. Li et al., "Drug specification named entity recognition base on BiLSTM-CRF model," in *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, pp. 429–433, Milwaukee, WI, USA, 2019.

[33] S. Gao, P. Jin, and D. Zhang, "Research on named entity recognition of TCM classics based on deep learning," *Technology Intelligence Engineering*, vol. 5, no. 1, pp. 113–123, 2019.