Named Entity Recognition Service of Bert-Transformer-CRF Based on Multi-feature Fusion for Chronic Disease Management

Diya Chen[1,2], Chen Liu[1,2](✉), and Zhuofeng Zhao[1,2](✉)

[1] School of Information, North China University of Technology, Beijing 100144, China
liuchen@ncut.edu.cn, edzhao@ncut.edu.cn

[2] Beijing Key Laboratory of Large-scale Stream Data Integration and Analysis Technology, Beijing 100144, China

Abstract. In chronic disease management, there are many kinds of chronic diseases, the professional terms are complex and diseases have their own characterised disease descriptions, making it difficult for some existing service recommendation methods to make accurate and personalised recommendations for patients with chronic diseases. Therefore, the paper proposes a service recommendation method based on the BERT-Transformer-CRF named entity recognition technology (BTC-SR) to achieve more accurate recommendation services. Firstly, the input disease text data is identified by a BERT (Bidirectional Encoder Representations from Transformers)-Transformer-CRF (Conditional Random Fields) model incorporating radical and pinyin features for named entities, then the relationships between the entities are extracted, and finally an implicit representation of the user is combined to present a recommendation list for the user. Experiments show that the proposed model achieves an F1 value of 60.15 for entity recognition on the CMeEE dataset, which provides better recognition results and lays the foundation for more accurate service recommendations.

Keywords: Chronic diseases · Named Entity Recognition Services · Multi-feature Fusion · Transformer

1 Introduction

With the rejuvenation of chronic diseases, chronic disease management is increasingly becoming a hot topic [1]. A vast amount of services are provided in the chronic disease management system. For example, intelligent interactive follow-up dialogue robots for multiple rounds of dialogue, patient education recommendations, condition assessment, follow-up services such as WeChat; personalised self-management tools for special patients or chronic disease populations,

The Key-Area Research and Development Program of Guangzhou City (202206030009).

Z. Wang et al. (Eds.): ICSS 2023, CCIS 1844, pp. 166–178, 2023.
https://doi.org/10.1007/978-981-99-4402-6_12

intelligent record indicators, medication recommendations; and a collaborative human-machine online communication and service platform. However, there are many types of chronic diseases and complex medical terminology involved, so how to provide accurate services for the chronic disease population is a research hotspot.

Service recommendation is a type of software service that understands user preferences through historical user-project interaction information and recommends services that may be of interest to users based on captured user preferences [2]. In recommendation algorithms, semantic-based recommendation algorithms require a lot of time and human resources in the case of complex service recommendations; keyword-based recommendation methods do not meet the personalised needs of users; content-based recommendation algorithms do not take into account the influence of long distances between texts when learning contextual information about users and are not applicable to service recommendations in the field of chronic diseases with complex terminology. In order to improve the effectiveness of service recommendation for chronic diseases, this paper proposes a service recommendation method (BTC-SR) based on the Bert-Transfer-CRF named entity recognition technique. This method first trains a word vector using Bert, combines the radical features and pinyin features, passes in the Transformer module to extract the features of each character's embedding, saves the flow of information over long distances between texts, decodes the predicted labels using the CRF layer, then extracts the entity-to-entity relationships, and finally combines the patient's implicit behaviour to recommend personalised services for them.

The BTC-SR model integration of radical features and pinyin features, improves the named entity recognition algorithm for long entities and specialized terms, setting epoch to 50, and achieves an F1 value of 60.15 on the CMeEE dataset, an improvement of 1.37 over the base model.

2 Related Work

This article uses Named entity recognition (NER) technology as the basis for service recommendation, so related work is done from recommendation algorithm and NER algorithm.

2.1 Algorithms for Service Recommendation

Recommendation algorithms are divided into traditional recommendation technologies and recommendation algorithms based on deep learning. Among them, traditional recommendation technologies include Semantic-based recommendation, keyword-based recommendation, content-based filtering, collaborative filtering, and hybrid recommendation [3]. Semantic-based recommendation algorithms, which require manual or semi-automatic construction of ontologies. Baidu Encyclopedia and Ding Xiang Yuan use search tools that use keyword-based matching methods that will return the same results to different users searching for the same keywords [4]. The content-based filtering algorithm takes the user's historical selection records or preference records as reference recommendations and mines unknown records and content that are highly relevant to the reference recommendations as system recommendations. Collaborative filtering algorithms obtain dependencies between users and items by analysing the user's rating matrix. Hybrid recommendation techniques incorporate different algorithms into the recommendation system. Nowadays, recommendation algorithms based on deep learning are widely used.

Content-based service recommendation algorithms, literature [5] proposed a service requirement discovery method based on matching the association rules between contextual information and services. The method obtains the association rules between context and service functions, and calculates the similarity of context information to construct a collection of personalised services available in the current application scenario. Based on collaborative filtering for service recommendation algorithms, the literature [6] proposes a cross-domain collaborative filtering model that extends user and item features through the potential factor space of auxiliary domains. The approach uses a Funk-SVD decomposition with an extended two-dimensional location feature vector and a C4.5 decision tree algorithm to predict missing ratings. The algorithm is based on hybrid recommendations, literature [7] proposed an optimal combination prediction idea based on collaborative filtering hybrid recommendation algorithm, which improves the accuracy of rating prediction and enhances the quality of recommendations.

Hansen et al. [8] proposed the CoSeRNN (Contextual and Sequential Recurrent Neural Network) model based on contextual modelling of user preferences. The model models user preferences as an embedding sequence for each conversation (session), and predicts user preferences by adding the user's historical behaviour and context at the beginning of the session.

2.2 Algorithms for NER

Early methods of NER are rule-based and dictionary-based, and later based on machine learning methods and deep learning-based methods. The first method is to develop rule templates by domain experts on the basis of existing knowledge and dictionaries to realize named entity recognition through matching [4]. Yuan J. et al. [6] integrated specific domain entity dictionaries, characteristic character rules and part-of-speech combination rules to identify named entities in the field of power safety, which proved its effectiveness on small sample data, and F1 reached 90. Machine learning-based NER technology uses the labeled training set, manually constructs features and labels words to achieve named entity recognition. Typical machine learning-based NER technologies include the Hidden Markov Model (HMM), the Maximum Entropy Markov Model (MEMM), the Support Vector Machine (SVM) model, and the Conditional Random Fields (CRF) model [7]. Zhai Juye et al. [9]on the basis of CRF, using language, keywords and dictionaries as features and optimizing them using rules. The accuracy, recall and F-value on Chinese electronic medical records have improved significantly, with accuracy increasing to 78.98 and recall and F-value improving to 88.37 and 83.41, which laid the foundation for subsequent NER research in the medical field.

NER technology based on deep learning can automatically discover hidden features in text, does not rely on manual feature selection, and dominates practical applications. Deep learning models mainly include convolutional neural networks (CNN) models, long short-term memory (LSTM) models, and bidirectional long short-term memory models. Sun Z., Li X. [11] after obtaining the Chinese character glyph features and labeling entity types, the fusion vector is sent to Mogrifier GRU, and finally decoded by CRF. It proves the effect of naming entity recognition in electronic medical records. Lian G. combined with the characteristics of network security domain entities, the structure of BI-LSTM-CRF model is improved, so that the model adapts to the recognition of named entities in the network security field, and the F value reaches 87, which provides ideas for the application of BI-LSTM-CRF model to other knowledge fields [12].

3 Service Recommendation Approach Based on NER

Chronic disease service recommendation is intended to provide patients with various chronic diseases with their personalized medication, diet, exercise method guide. Taking hypertension as an example, if the patient is less than 65 years old, simple diastolic hypertension, heart rate is not more than 80 beats per minute, ACEI antihypertensive drugs will be recommended for him, and if it is isolated systolic hypertension in the elderly, CCB is recommended. In terms of diet, patients are recommended with suitable and fasted foods. Before these service recommendations can be provided, firstly, named entities need to be extracted from the disease text, extracting disease symptom entities, drug entities and food entities etc. Secondly, relationships between entities are established so that they can be matched with the user's personalised profile to complete service recommendations. The technical roadmap and NER technology model architecture BERT-Transformer-CRF proposed in this paper are shown as Fig. 1.

The NER technology model architecture BERT-Transformer-CRF proposed in this paper is shown in Fig. 2. First, the input entity features are extracted by fine-tuning the pre-training of the BERT model to fuse the word vector representation sequence with the radical features and the pinyin features. Then, dependency features between long-range texts are obtained by Transformer and features of long entities in the text are learnt. Finally, contextual annotation constraints are learnt in the CRF module and the output entity labels are decoded.

3.1 Transformer Layer

Transformer is a model based entirely on the attention mechanism to improve the speed of model training, using the encoder-decoder architecture. The encoder converts the input sequence $(x_1, ...x_n)$ into a continuous expression $(z_1, ...z_n)$, and the decoder generates the output sequence $(y_1, ...y_n)$ based on that expression. Scaled Dot-Product Attention is an important component of multi-headed attention, which is essentially an attention mechanism that uses dot product

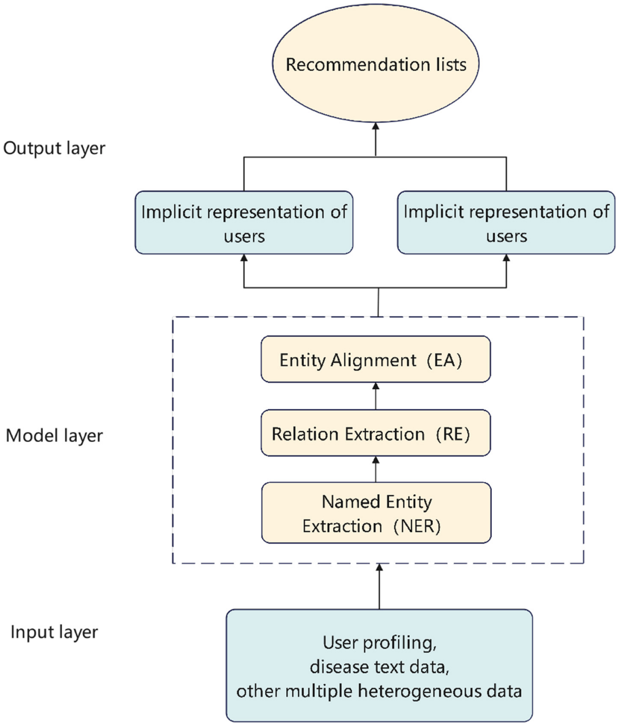

Fig. 1. Process architecture diagram for service recommendations

design to calculate similarity. Figure 3 shows the computation flow of the scaled dot product attention, which is shown in Eq. (1).

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{1}$$

where: $\sqrt{d_k}$ is the dimension of query vector and key vector, and it is the penalty term.

Multi-head attention is composed of multiple self-attention stitching, and the working principle is as follows: After linear transformation of Query and Key-Value, the similarity is calculated using Scaled Dot-Product Attention, and the same operation is performed h times. "h" is the number of layers of multi-headed attention. Finally, the results of each layer are stitched together to obtain feature information from different angles and at different levels. Figure 4 shows the detailed flow of Transformer's multi-headed attention mechanism, and the corresponding calculation formula is shown in Eq. (2) (3). The number of self-attention of the transformer is the number of heads, and each head focuses on different contextual information for each word.

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{2}$$

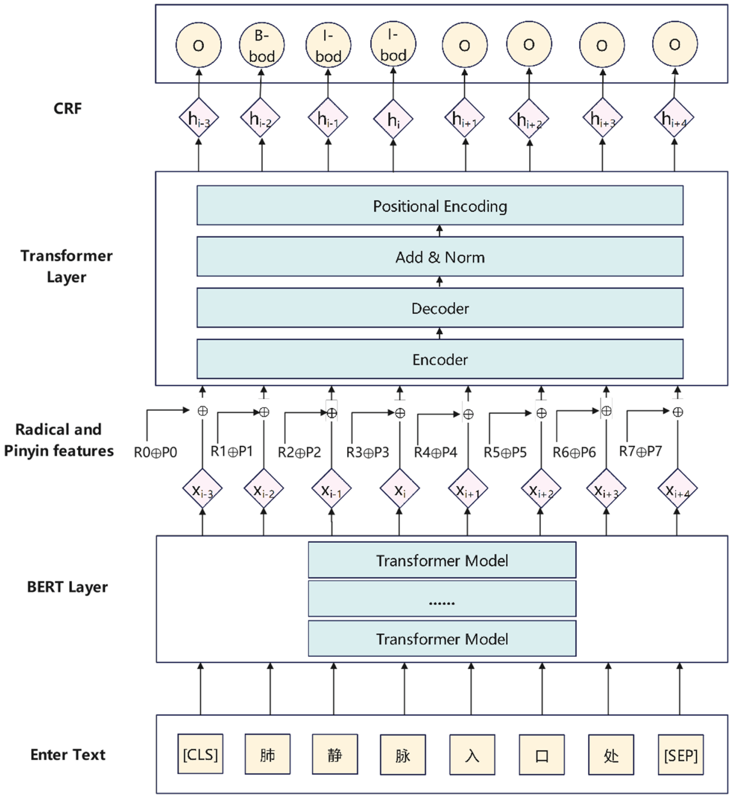

Fig. 2. Model for BERT-Transformer-CRF+radical+pinyin

where: W^Q, W^K, W^V is weight matrixs.

The output matrix of multi-head attention is as follow

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W^O \tag{3}$$

where: Concat means that the results of each layer of the scaled dot-product attention are stitched together.

3.2 CRF Layer

CRF is a basic model for natural language processing and is widely used in scenarios such as word segmentation, entity recognition, and part-of-speech tagging. Through CRF, the model can automatically learn certain constraints to ensure the legitimacy of the predicted label. Secondly, CRF has a transfer feature, considering the order between labels, and automatically learning the constraints of sentences.For example,the entity's start label should be "B-dis", not "I-dis". Therefore, introducing CRF to solve Transformer only considers long distance

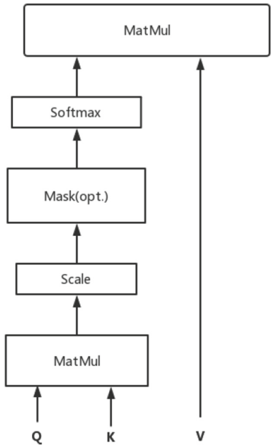

Fig. 3. Scaled Dot-Product Attention

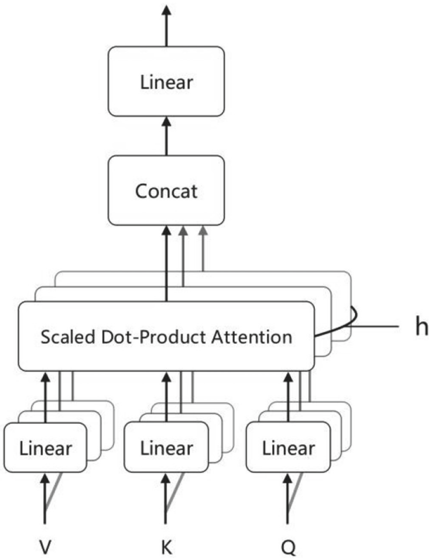

Fig. 4. Multi-head attention structure

dependencies between contexts. For a given input sequence $X = (x_1, ...x_n)$, the sequence probability score predicted by CRF is shown in the following Eq. (4). The normalization formula by the Softmax function is shown in the following Eq. (5) to obtain the probability values of each position of the predicted sequence.

$$score(X, y) = \sum_{i=0}^{n} T_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{i, y_i} \tag{4}$$

$$p(y|X) = \frac{e^{s(X,y)}}{\sum_{\tilde{y} \in Y_X} e^{s(X,\tilde{y})}} \tag{5}$$

$$Loss = log(p(y|X)) \tag{6}$$

where: $\tilde{y}$ is the real label; Y_X is given disease text all possible label combinations; *score* is the score of the degree of correspondence between the defect text y and the X tag sequence; *Loss* is the loss function.

When the final prediction of the text entity label of medical disease is made by CRF, the global optimal solution is obtained by using the viterbi algorithm, and the solution formula is

$$y^* = argmaxscore(X, \tilde{y}) \tag{7}$$

where: y^* is the sequence of medical disease text labels with the highest score.

4 Experiment and Result

4.1 Dataset Labels

The essence of the NER service based on deep learning is a sequence labeling task, that is, to identify the entity information with specific meaning in the sentence. The following will select the appropriate dataset labeling system to complete the dataset labeling.

Dataset Labels. There are three common methods for sequence annotation, namely BIO, BMES, and BIOSE. BIO annotates the beginning, end and non-entity parts of the entity; BMES, add S as the annotation of a single word entity; The BIOS adds E to the end label of the entity. In this example, the BIO annotation mode is selected. The identification of named entities in the medical field includes 9 types of entities, namely disease names, clinical manifestations, medical procedures, medical equipment, drugs, medical test items, bodies, departments, and microorganisms. The definitions of each category tag are shown in the Table 1, and each label is an abbreviation of the meaning of the entity.

Table 1. Label representations of each entity category

Number	Entity Class	Begin Label	Middle Label	End Label
1	Disease name	B-dis	I-dis	I-dis
2	Symptom	B-sym	I-sym	I-sym
3	Procedure	B-pro	I-pro	I-pro
4	Medical-equipment	B-equ	I-equ	I-equ
5	Drug	B-dru	I-dru	I-dru
6	Medical test item	B-ite	I-ite	I-ite
7	Body	B-bod	I-bod	I-bod
8	Department	B-dep	I-dep	I-dep
9	Microorganism	B-mic	I-mic	I-mic

Dataset Introduction. The data extracted for named entities in this paper is the Chinese medical naming entity recognition data CMeEE in the dataset combination of the previous academic evaluation competitions of CHIP Conference and Aliquark medical search business in 2021, which is rich in corpus and adds medical dialogues, electronic medical records and medical imaging reports written by medical experts, which is more time-sensitive and professional, and the data covers a wide range, and the use of this dataset for named entity recognition research is highly representative.

The dataset includes 15000 pieces of data from the training set, divide the test set by an 8:2,5,000 pieces of data from the validation set, and 3,000 pieces of data from the test set. The total number of words in the annotation data reached 2.2 million, containing 47,194 sentences and 938 files, with an average word count of 2,355 per file. The dataset contains 504 common pediatric diseases, 7,085 body parts, 12,907 clinical manifestations, and 4,354 medical procedures.Table 2 lists the different categories of entity statistics.

Table 2. Statistics of different types of medical entities on CMeEE dataset

Data Set	Training Data	Test Set
Disease name	12673	3170
Symptom	9730	2539
Procedure	4994	1338
Medical-equipment	695	193
Drug	3122	808
Medical test item	2059	522
Body	14178	3519
Department	271	77
Microorganism	1557	351

4.2 Experimental Environment and Evaluation Indicators

The following is an introduction to the experimental environment and evaluation indicators.

Experimental Environment and Parameter Settings. The experiment was carried out in the Linux operating system, developed in Python 3.6 language, the deep learning framework Tensorllow version is 1.15.0, the Bert model is the Bert-Base-Chinese version, and the hardware configuration: CPU model is i5-10400 2.90 GHz, GPU model is GTX2080. The important parameter settings of the model are shown in the Table 3.

Table 3. Model parameter settings

Parameter	Value
batch size	64
dropout rate	0.5
learning rate	1e−05
maxseq length	202
epochs	50

Evaluation Indicators. Evaluate with accuracy, recall, and F1 values. Accuracy is how many of the samples that are predicted to be correct are truly correct; Recall is how many of all the correct samples are correctly identified as positive; The F1 value is the weighted harmonic average of accuracy P and recall R. The higher the evaluation metric data, the better the performance of the entity extraction model. Its calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

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

4.3 Comparison of Experiment Results

In order to verify the validity of the proposed method, the following five main named entity recognition algorithms are selected for comparison. Including machine learning algorithm CRF, deep learning and machine learning fusion algorithm BERT-BI-LSTM-CRF, the experimental results are shown in Table 4.

The comparison of CRF and BERT-BI-LSTM-CRF shows the effectiveness of deep learning algorithm, which the P, R, F1 values are increased by 0.71, 2.96, 0.46. The comparison of BERT-BI-LSTM-CRF and BTC shows that Transformer can obtain better feature representation, and the P, R and F1 values are increased by 4.09, 6.31 and 3.57 respectively, which proved the effectiveness of Transformer and was better than other algorithms. It is fully proved that Multi-head attention in Transformer has stronger long-distance dependency representation capabilities. Moreover, By applying pinyin and radical features to the fine-tuned BERT model, the Transformer can obtain richer semantic information, and F1 is improved by 1.37. In addition, the P, R, and F1 of each type of entity are compared. Department and Medical test item recognition is poor, in the Disease name, Drug, Microorganism achieved good results. For example, in disease name, terminology such as "horseshoekidney", "unilateralfusion", "humandiploidcell" are correctly predicted. In drug, predictions for long entities such as "Golden Hamster Kidney Cell (GHKC) and Meriones Gerbil Kidney Cell (MGKC) Inactivated Vaccine" are accurate, indicating that Transformer has a good effect on the identification of long entities. As shown in Table 5.

Table 4. Model parameter settings

Model	P	R	F1
CRF	52.41	54.24	54.75
BERT-BI-LSTM-CRF	53.14	57.20	55.21
BTC	57.23	63.51	58.78
BTC+radical+pinyin	58.43	64.82	60.15

Table 5. Recognition and comparison of different types of entities on CMeEE

Entity Type	P	R	F1
Disease name	69.71	71.7	70.69
Symptom	40.44	46.37	43.21
Procedure	46.05	48.93	47.45
Medical-equipment	53.22	66.31	59.05
Drug	68.28	70.25	69.25
Medical test item	31.41	36.57	33.79
Body	51.97	66.56	58.36
Department	15.97	21.22	16.99
Microorganism	71.73	78.72	75.06

5 Conclusion

Based on personalized recommendation services in the field of chronic diseases, this paper proposes a service recommendation method based on multi-feature fusion Bert-Transformer-CRF named entity recognition technology. First, Bert is used to train word vectors to obtain contextual features. After adding radicals and pinyin features, it is passed into Transformer module to extract embedded features of each character and save remote information flow between texts. The CRF layer decodes the prediction label to obtain the entity classification label. Then the relationship between entities is extracted and the implicit behavior of users is combined with the recommendation. Experimental results show that this method has better named entity recognition effect, which proves its availability in service recommendation, and lays a more solid foundation for improving the accuracy of personalized recommendation service.

References

1. Li, R., Shi, B., Guo, H., Liu, F., Li, T., Li, M.: Research progress of chronic disease management in internet hospitals. Chin. Math. **18**(01), 95–101 (2023)
2. Zhang, M., Zhang, X., Liu, S., Tian, H., Yang, Q.: Review of recommendation systems using knowledge graph. Comput. Eng. Appl. **59**(04), 30–42 (2023)
3. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems a survey of the state of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. **17**(6), 734–749 (2005)
4. Li, Y., Wang, R., Nan, G., Li, D., Li, M.: A personalized paper recommendation method considering diverse user preferences. Decis. Support Syst. **12**(07), 49–56 (2021)
5. Zhang, Z., Zhang, X.: Context-aware information service recommendation method for high-speed rail. Comput. Eng. Appl. **57**(12), 231–236 (2021)
6. Yu, X., Jiang, F., Du, J., Gong, D.: A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains. Pattern Recognit. **94**, 96–109 (2019)
7. Yu Q., Zhao M., Luo Y.: Collaborative filtering hybrid recommendation algorithm based on optimal weight and its application. Oper. Res. Manage. Sci. https://kns.cnki.net/kcms/detail//34.1133.G3.20230227.1815.006.html
8. Hansem, C., Maystre, L.: Contextual and sequential user embeddings for large-scale music recommendation. In: Proceedings of the 14th ACM Conference on Recommender Systems, pp. 53–62. ACM Press, New York (2020)
9. Li, D., Luo, S., Zhang, X., Xu, F.: A review of named entity recognition methods. J. Front. Comput. Sci. Technol. **16**(9) (2022)
10. Yuan, J., Pan, M., Zhang, T., Jiang, Y.: Named entity recognition in the field of power safety based on rules and dictionaries. Electron. Technol. Appl. **48**(12) (2022)
11. Jiang, Q., Gui, Q., Wang, L.: Review of research progress of named entity recognition technology. Electr. Power Inf. Commun. Technol. **20**(2), 15–24 (2022)
12. Zhai, J., Chen, C., Zhang, Y.: Research on entity recognition of Chinese electronic medical record naming based on the combination of CRF and rules. J. Baotou Med. Coll. **33**(11) (2017)

13. Sun, Z., Li, X.L.: Multi-feature fusion Chinese electronic medical record named entity recognition. Comput. Eng. https://kns.cnkinet/kcms/detai1//11.2127.TP20221213.1206.005.html
14. Lian G.: Construction of knowledge graph in network security domain based on improved bi-LSTM-CRF. Xinxijishu **12**(2), 130–142 (2022)