An Autonomous Dual-channel Entity Recognition Method for Chinese Acupuncture

Anonymous Author(s)
Affiliation
Address
email

Abstract

Due to the strong specialization of knowledge in the field of acupuncture and moxibustion, the entities are more closely related, the structure is complex, and there are more mixed words in Chinese and English, which leads to the poor effect of the general entity recognition model in the field of acupuncture and moxibustion. In this regard, a two-way meridian entity recognition model is proposed: for the entity categories of acupoints, manipulation techniques, and evidence types that are more fixed and contain corresponding keywords, we adopt the incorporation of lexicon information into the bottom layer of BERT to enhance the effective utilization of lexicon information, and we introduce a contribution factor in BiLSTM to improve the correlation before and after the statements and enhance the semantic information; for the diseases and symptoms that have large vocabularies and are characterized by a large amount of abbreviations, an extra attention mechanism is introduced in BERT to obtain word-level features and semantic information in different dimensions, to improve the shortcomings of the deep semantic information output from BERT which lacks the underlying word-level features; finally, an integration of the results of the two-way warping is done. In order to verify the effectiveness of the proposed model, it is compared with the existing methods on the labeled dataset, and the F1 value of the model on the dataset can reach 94.21percent. The experimental results prove that the proposed dual-road meridian entity recognition model can recognize entities in the field of acupuncture and moxibustion better than other methods.

1 Introduction

Acupuncture and moxibustion, after thousands of years of historical development, has accumulated a huge amount of scientific and cultural knowledge in theoretical research and countless clinical practices, and most of this knowledge is stored in the ancient books of traditional Chinese medicine, scientific literature, and clinical cases in hospitals, which is of great value and significance for research[1]. However, because these resources are scattered, most of them are presented in the form of text, resulting in less explicit knowledge and noisy data, and the related medical case information is also fragmented and fragmented, so it will make the inheritance of acupuncture and moxibustion have certain obstacles[2], so there is an urgent need to build a knowledge map of acupuncture and moxibustion, to deeply excavate the potential tacit knowledge, and to build a relatively complete knowledge map[3].

As the core of knowledge graph construction, entity recognition aims to accurately extract key information from complex unstructured and semi-structured texts. So far, entity recognition methods can be mainly divided into three types: rule-based and dictionary-based methods, statistical machine learning based methods, and deep learning based methods.

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

Early methods: Rule-based methods usually involve linguistic experts manually constructing rule templates based on language characteristics, and then using matching to achieve entity recognition. The dictionary-based method requires first constructing an entity dictionary, and then identifying entities by searching the dictionary[4]. This method is effective in specific fields, but it has high costs and poor portability, making it difficult to be widely applied across different domains.

Statistical machine learning: To solve the above problems, researchers turn to machine learning techniques such as Hidden Markov Models (HMM)[5], Maximum Entropy Markov Model (MEMM) [6] and Conditional Random Fields (CRF)[7]. These methods treat entity recognition as a sequence labeling task, where CRF effectively solves the label bias problem of MEMM through global normalization, demonstrating its advantages in long-distance dependencies and domain information fusion. However, these methods still require a large amount of manually annotated corpus and are expensive.

Deep learning: With the rise of deep learning technology, models based on convolutional neural networks (CNN)[8], recurrent neural networks (RNN)[9], long short-term memory networks (LSTM) [10] and bidirectional LSTM (BiLSTM)[11] have been applied to entity recognition. CNN improves efficiency by reducing feature engineering, but it is prone to lose global features; RNN and its variants LSTM, BiLSTM effectively capture the long-term dependencies in sequences. The BiLSTM-CRF[12] model solves the problem of unreasonable label prediction by combining with a CRF decoder. To deal with complex situations, researchers have also introduced strategies such as attention mechanisms and multi-layer CNNs to further improve model performance.

Pre-trained models: In recent years, the excellent performance of pre-trained models such as BERT[13] in natural language processing has promoted their application in entity recognition. Models such as BERT-CRF[14] have significantly improved the accuracy of entity recognition by combining pre-trained language representations with sequence labeling techniques, especially when dealing with scarce labeled data. The entity recognition method based on deep learning can perform well in the task of entity recognition by constructing a corresponding neural network model[15].

Since most of the classical entity recognition models are proposed in the general domain research, and the research in specific domains is only limited to military, agriculture, maritime, Chinese medicine prescription, etc., 6there is a certain gap in the research related to the entity recognition of Chinese medicine and acupuncture. And the main difficulties of entity recognition in the field of Chinese medicine and acupuncture compared to the traditional classical entity recognition model research are: (1) the specificity of the field requires the acquisition of high-quality data, and the specificity of the field leads to the acquisition of less textual information, which affects the accuracy rate of the model; (2) the field has more terminology, which leads to the poorer generalization of the traditional model, which affects the accuracy rate of the entity recognition; (3) the domain text entities are complex in composition, contain a large number of mixed words such as numbers and letters as well as close relationships within the entities, and it is difficult for traditional entity recognition models to learn such complex text relationships; (4) the number of entity types in the field of acupuncture and moxibustion is not balanced, and different entity types also have their own characteristics, which leads to a lower accuracy rate of recognition by traditional entity recognition models. Based on this this paper proposes a two-path acupuncture domain entity recognition method based on improved BERT-BiLSTM-CRF to improve the accuracy of entity recognition.

2 Methodology

Text data in the field of acupuncture and moxibustion poses unique challenges, including the prevalent use of clinical abbreviations for diseases and symptoms, excessively long entity recognitions, polysemy (multiple meanings for a single word), and complex entity boundaries. For instance, "headache" can be classified as both a "symptom" and a "disease" depending on the context. To address these issues, we propose an enhanced dual-path BERT-BiLSTM-CRF entity recognition method.

We leverage the pre-trained language model BERT to extract global feature information from text data in the acupuncture and moxibustion domain. This approach can effectively mitigate issues such as polysemy, clinical abbreviations, and lengthy entities to a certain extent. In BERT's architecture, the encoding layers are composed of multiple Transformer layers. While encoders closer to the top layers capture deeper semantic information, they may lack low-level lexical features. Conversely, shallow encoders excel at lexical features but lack deep semantic insights. Since BERT only outputs the result

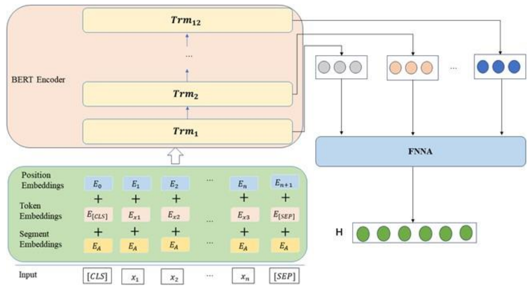

Figure 1: Structure Diagram of MLFBERT

from the top encoder layer, it can inadvertently lose lexical information, limiting the performance of the raw BERT model when directly applied to binary classification datasets in the acupuncture and moxibustion field.

To enhance the model's performance for binary classification tasks, we introduce an attention mechanism at the bottom of BERT. This mechanism assigns a weight score to the output of each encoding layer in BERT. By aggregating the information from each layer through weighted outputs, the attention mechanism prioritizes important layers with higher weights and downplays less significant layers, thus addressing the limitation of the original BERT model's exclusive reliance on high-level semantic information and lack of low-level lexical features.

The green section represents the input layer of BERT, which encodes the input sequence. To address issues such as polysemy and excessively long words within sentences, the BERT model embeds the text into three components: word embeddings, sentence-level embeddings, and positional embeddings, resulting in the final word vector input. The calculation formula for this is shown below.

$$V = E_s + E_t + E_p$$

Where E_s, E_t, $E_p$respectively represent sentence-level embeddings, word embeddings, and positional embeddings; the blue section depicts the Forward Neural Network Attention Mechanism (FNNA).

Here, our MLFBERT model is derived by incorporating the FNNA (Feed-forward Neural Network Attention) mechanism into the encoding layers of BERT. FNNA is a feed-forward neural network attention mechanism that learns the relationships between the output features of different encoding layers and adaptively adjusts the weights of these features based on the task. This approach reduces the loss of important information during information transmission and enables the model to capture multi-level textual information, thereby enhancing model performance.

To further enhance the model's ability to understand deep contextual information, we incorporate the BiLSTM model after MLFBERT. This is because BiLSTM is better at capturing long-distance dependencies within sequences. Combining it with the MLFBERT model can leverage the strengths of both, thereby improving the model's performance in entity recognition tasks.

Finally, the CRF model is chosen as the decoder for the binary classification entity recognition task due to the limitations of the commonly used Softmax function as the activation function in the output layer. Softmax merely outputs based on prediction probabilities without considering the relationship between output labels and contextual labels, leading to isolated entity labels and the model falling into local optima. However, in sequence labeling tasks, there is an inherent correlation between subsequent entity labels. For instance, the label "B-Acu" is highly likely to be followed by "I-Acu" or "O", but not "I-OperationMethod". If "I-OperationMethod" appears, it indicates an incorrect entity prediction. Therefore, the CRF model is selected as the decoder, as it can take a global perspective, avoid outputting illogical label sequences, and enhance the recognition capability of the binary classification entity recognition model.

In the case of three-class classification datasets, entities often exhibit strong correlations with their preceding and succeeding words. For instance, when encountering terms like "Hegu Acupoint", "Damp-heat Syndrome", and "Lifting and Thrusting Method", the presence of "Acupoint", "Syndrome", and "Method" can greatly suggest that "Gu", "Heat", and "Thrusting" are also parts of the

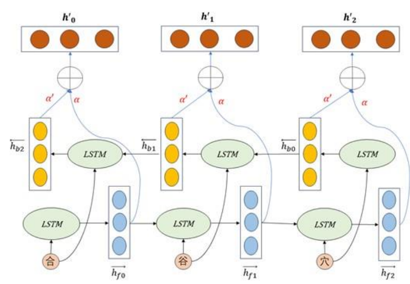

Figure 2: Structure Diagram of BiLSTMWW

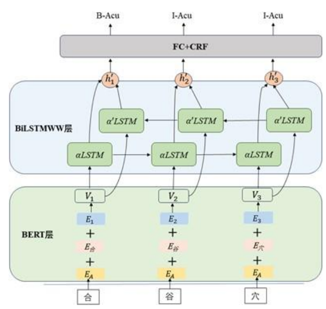

Figure 3: Structure Diagram of BERT-BiLSTMWW-CRF

corresponding entities. To strengthen the connection between forward and backward propagation in BiLSTM, we refine the fusion method of BiLSTM’s forward and backward passes. This allows for a selective output based on the relative contributions of both directions. We introduce a learnable weight to adjust the fused output of sequential forward and backward information, resulting in an improved BiLSTM model known as BiLSTMWW (Bidirectional Long Short-Term Memory With Weighted) model. The model architecture is illustrated in the following diagram.

We multiply the vectors obtained from the forward and backward passes by their respective weight factors&’ and and sum them to derive two feature vectors, which are then combined to produce the output vector for the current character encoded by BiLSTMWW. Here, the dynamic weight, denoted as , is determined through extensive experimentation with =1. Based on this design, we integrate BERT and CRF to realize the BERT-BiLSTMWW-CRF model for the task of tri-class entity recognition. The overall model architecture diagram is presented below.

Based on the above design, the BERT-BiLSTMWW-CRF model is implemented by integrating BERT and CRF for the ternary classification entity recognition task. The overall model structure diagram is shown below.

After validating the performance of the aforementioned models, we integrate the binary classification entity recognition model with the tri-class classification entity recognition model. Through these two parallel paths, we respectively perform the binary and tri-class entity recognition tasks. Finally, the results from both paths are fused according to the established rules to yield the final five-class entity recognition results. The structure diagram of the improved BERT-BiLSTM-CRF dual-path entity recognition model is shown below.

The fusion rules adopted here are as follows: (1) The model MLFBERT-BiLSTM-CRF is designated as Model 1, and the model BERT-BiLSTMWW-CRF is designated as Model 2. (2) If both Model 1 and Model 2 predict a certain character as "O", then the entity type corresponding to that character is "O". (3) If neither Model 1 nor Model 2 predicts a certain character as "O", and there is a conflict in the entity types predicted by the two models for that character, then the entity type of that character is set to "O". (4) If Model 1 predicts a certain character as "O", but Model 2 predicts it as not "O", then the entity type of that character is the prediction result of Model 2. (5) If Model 2 predicts a certain character as "O", but Model 1 predicts it as not "O", then the entity type of that character is the prediction result of Model 1.

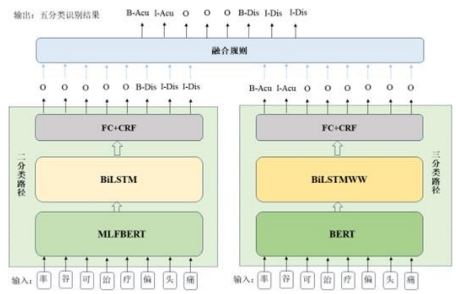

Figure 4: Improved BERT-BiLSTM-CRF dual path entity classification model

Table 1: Parameter configuration table for binary classification experiments

Experimental parameters	parameter value
Number of BERT layers	12
BERT dimension	768
learning rate	0.00001
Dropout	0.5
optimizer	Adam
LSTM dimension	128
Maximum sequence length	150
Batch_size	32

3 Experiments and Results

In this paper, the experimental results and analysis of the proposed improved BERT-BiLSTM-CRF based dual path entity recognition method are presented.

3.1 Model Parameters

In this paper, two kinds of experiments are conducted, which are two-classified entity recognition and three-classified entity recognition. And the experiments are carried out using MLFBERT-BiLSTM-CRF binary classification entity recognition method and BERT-BiLSTMWW-CRF triclassification entity recognition method.The MLFBERT-BiLSTM-CRF model parameters are shown in Table 1, and the BERT-BiLSTMWW-CRF model parameters are shown in Table 2.

3.2 Experimental data and evaluation indexes

The raw data of this paper are mainly taken from the text data in journals and guidelines of the three types of diseases: incontinence, pain and nerve injury, in which the journals and guidelines are determined by specialist professor of acupuncture and moxibustion, which ensures the accuracy, authority, credibility and feasibility of the data to a certain extent. The dataset contains a total of 1747 documents and 5 guideline specifications. The number of entities in the dichotomous dataset, the trichotomous dataset and the pentachotomous dataset are shown in Table 3.

Table 2: Parameter Configuration Table for Three Classification Experiments

Experimental parameters	parameter value
Number of BERT layers	12
BERT dimension	768
learning rate	0.00001
Dropout	0.1
optimizer	Adam
LSTM dimension	128
Maximum sequence length	150
Batch_size	32

Table 3: Number of entities

Type of data set	entities in the training set	validation set entities	test set entities	total
dichotomous data set	28120	8244	4005	40369
Triple categorical datasets	24106	6507	3588	34201
Penta-classified data sets	52226	14751	7593	74570

Table 4: Comparative experiments on different entity recognition methods

Number	model	P(%)	R(%)	F(%)
1	LSTM	68.89	57.31	62.57
2	BiLSTM	73.72	74.12	73.92
3	BiLSTM-CRF	78.79	76.15	77.45
4	BERT-CRF	88.66	90.72	89.68
5	BERT-BiLSTM-CRF	89.21	90.87	90.03
6	MLFBERT-BiLSTM-CRF	90.88	93.21	92.03

Facing the entity recognition in the field of acupuncture and moxibustion, the two main perspectives of checking the full rate and checking the accuracy rate are considered, and thus Precision (P) and Recall (R) are adopted as the indicators for evaluating the model, and in order to comprehensively assess the performance of the model, is also introduced to comprehensively evaluate the P and R indicators. In previous studies P and R usually account for equal ratios, so=1 is commonly defaulted, and is uniformly denoted as F1 in the subsequent part of this paper.

3.3 Results and Analysis

In order to validate the effectiveness of the proposed dual-path entity recognition method, experiments were conducted with MLFBERT-BiLSTM-CRF on the two-classified dataset, LEBERT-BiLSTMWW-CRF on the three-classified dataset, and the fusion of the dual-path model results on the five-classified dataset, respectively.

3.3.1 MLFBERT-BiLSTM-CRF binary classification entity recognition model experiments

Two kinds of experiments were designed, the comparison experiment of different entity types and the comparison experiment of different entity methods. In the comparison experiments of different entity types, the model is compared in the experiments of two types of entities, disease and symptom, in which the recognition effect is better for the disease entity types, mainly because the number of disease entity types in the training data is more as well as the name is more fixed.

In the comparison experiments of different entity recognition methods, the MLFBERT-BiLSTM-CRF model is compared with the performance of five models, namely, LSTM, BiLSTM, BiLSTM-CRF, BERT-CRF, and BERT-BiLSTM-CRF, respectively, and the comparison results are shown in Table 4.

Through comparison experiments, it is found that the MLFBERT-BiLSTM-CRF proposed in this paper achieves the best results in all evaluation metrics. Through experiment 6, it is found that weighted fusion output of the output features of the 12 coding layers of the BERT model through FNNA after assigning different attention scores to the features of different layers makes the model improve greatly in all indicators, using BiLSTM can consider the contextual information of the sentence at the same time, which can capture the information in the sequence in a better way, and extract a better representation of the features, and the CRF model is able to solve the problem that the output of the neural network model does not conform to the common sense labeling, thus maximizing the performance of the model.

3.3.2 Validation Experiments of BERT-BiLSTMWW-CRF Triple Classification Entity Recognition Model

In the comparison experiment of different entity types, the model is compared in three types of entity types: acupoints, evidence types and manipulation techniques. The model is the best for the recognition of acupoint entities because there is more training data for the acupoint entity type and the model can easily learn its features.

Table 5: Comparative experiments on different entity recognition methods

Numb r	model	P	R	F
1	BiLSTM-CRF	76.32%	59.02%	66.56%
2	BERT-CRF	94.77%	95.03%	94.90%
3	BERT-BiLSTM-CRF	95.19%	96.11%	95.64%
4	BERT-BiLSTMWW-CRF	96.27%	96.99%	96.63%

Table 6: Comparative experiments on different entity recognition methods

Number	model	P	R	F1
1	BiLSTM-CRF	83%	83.80%	83.40%
2	BERT-CRF	91.46%	93.58%	92.50%
3	BERT-BiLSTM-CRF	90.67%	93.67%	92.15%
4	BERT-BiLSTMWW-CRF	92.94%	94.83%	93.87%
5	MLFBERT-BiLSTM-CRF	93.32%	94.58%	93.95%
6	Improved BERT-BiLSTM-CRF Dual Path	93.46%	94.97%	94.21%

In the comparison experiments of different entity recognition methods, the BERT-BiLSTMWW-CRF model is compared with three models, BiLSTM-CRF, BERT-CRF, and BERT-BiLSTM-CRF, respectively, to validate the effectiveness of the model under other conditions, and the results of the comparison are shown in Table 5.

Comparative experiments reveal that the BERT-BiLSTMWW-CRF model achieves the best results in all metrics. The introduction of the weighting parameter BiLSTM in this model has led to an improvement in the entity recognition performance of the model, with an increase of 1.08% in precision, 0.88% in recall, and 1.01% in F1 value compared to BERT-BiLSTM-CRF. This is due to the fact that the weight parameter can be selectively output according to the degree of contribution of forward and backward propagation, so that the two processes of forward and backward propagation can obtain better results by constraining each other.

3.3.3 Experiments on the improved BERT-BiLSTM-CRF dual-path five-classification entity recognition model

Based on the fusion rule the results of two-classified entity recognition and three-classified entity recognition are fused to get the final five-classified entity recognition results. Two kinds of experiments are designed, the comparison experiment of different entity types and the comparison experiment of different entity methods. In the comparison experiment of different entity types, the proposed two-path entity recognition model based on the improved BERT-BiLSTM-CRF is compared experimentally among the five categories of entity types, namely, disease, symptom, acupoint, evidence and manipulation. The F1 value of the model for recognizing the three types of entities, namely, acupoints, diseases, and evidence types, is relatively high, because these types of entities have more training data, and the model is easy to learn their features, and on the whole the model is able to achieve a good result for entity recognition of knowledge in the field of acupuncture and moxibustion.

In the comparison experiments of different entity methods, the two-path entity recognition model based on the improved BERT-BiLSTM-CRF is compared with five models, BiLSTM-CRF, BERT-CRF, BERT-BiLSTM-CRF, MLFBERT-BiLSTM-CRF, and BERT-BiLSTMWW-CRF, respectively, and under the All other conditions are the same to verify the validity of the model, and the experimental results are shown in Table 6.

Comparative experiments reveal that the improved BERT-BiLSTM-CRF based dual path entity recognition model achieves the best results in all evaluation metrics. The model is improved in all indicators compared to BERT-BiLSTM-CRF, the precision rate P is improved by 2.78%, the recall rate R is improved by 1.31%, and the F1 value is improved by 2.05%; and the precision rate P is improved by 10.46%, the recall rate R is improved by 11.17%, and the F1 value is improved by 10.81% compared to the BiLSTM-CRF model. This is because the proposed dual-path model fully considers the characteristics of different entity type datasets and proposes dual-path models for different features, which makes the model on each branch to learn fewer features thus generating high performance, and improves the model on each branch by introducing contributing factors to the three-classified entities to strengthen the linkage between the forward propagation and the backward propagation of the BiLSTM, and to obtain more global feature information; the way of introducing the attention mechanism for the two-classified entities to acquire the features of different levels of BERT, which strengthens the connection between the words and the semantic information and makes the model learn more feature information; thus giving good results for the overall model performance.

4 Conclusion

This paper mainly describes the two-classified entity recognition method and three-classified entity recognition method included in the improved BERT-BiLSTM-CRF two-path entity recognition method, and then proposes a result fusion rule for the two-classified entity recognition results and the three-classified entity recognition results, according to which the results in the two-paths are fused to obtain the final five-classified entity recognition results; at the same time, the two-classified entity recognition method and the five-classified entity recognition method proposed in this chapter are compared with other methods in different aspects, and the comparison results are given. At the same time, the two-classified entity recognition method, three-classified entity recognition method and five-classified entity recognition method with two-path result fusion proposed in this paper are compared with other methods in different aspects, and the comparison results are given. Through the analysis of the experimental results, it can be seen that the improved BERT-BiLSTM-CRF dual-path acupuncture entity recognition method proposed in this paper is the most effective.

5 References

1. Han X, Li X, Liang Y, et al. Acupuncture and Tuina Knowledge Graph for Ancient Literature of Traditional Chinese Medicine[C]. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2021: 674-677.
2. Bikel D M , Schwartz R , Weischedel R M . An Algorithm that Learns What's in a Name[J]. Machine Learning, 1999, 34:211-231.
3. Zhou G D, Su J. Named entity recognition using an HMM-based chunk tagger[C]. Proceedings of the 40th annual meeting of the association for computational linguistics. 2002: 473-480.
4. Strubell E , Verga P , Belanger D , et al. Fast and Accurate Entity Recognition with Iterated Dilated Convolutions[C]. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017: 2670-2680.
5. Eddy S R. Hidden markov models[J]. Current opinion in structural biology, 1996, 6(3):361-365.
6. McCallum A, Freitag D, Pereira F C N. Maximum Entropy Markov Models for Information Extraction and Segmentation[C]. ICML. 2000, 17(2000): 591-598.
7. Sutton C, McCallum A. An introduction to conditional random fields[J]. Foundations and Trends® in Machine Learning, 2012, 4(4): 267-373.
8. Collobert R , Weston J , Bottou L , et al. Natural Language Processing (almost) from Scratch[J]. Journal of Machine Learning Research, 2011, 12(1):2493-2537.
9. Goller C , Kuchler A . Learning task-dependent distributed representations by backpropagation through structure[C]. Proceedings of International Conference on Neural Networks (ICNN'96). 1996:347-352.
10. Gers F A, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM[J]. Neural computation, 2000, 12(10): 2451-2471.
11. Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks[C]. 2013 IEEE international conference on acoustics, speech and signal processing, 2013: 6645-6649.
12. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging[J]. arxiv preprint arxiv:1508.01991, 2015.
13. Zhou P, Shi W, Tian J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2016: 207-212.
14. Souza F, Nogueira R, Lotufo R. Portuguese named entity recognition using BERT-CRF[J]. arxiv preprint arxiv:1909.10649, 2019.
15. Kong J, Zhang L, Jiang M, et al. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition[J]. Journal of Biomedical Informatics, 2021, 116:103737.