Original Paper

Chinese Clinical Named Entity Recognition From Electronic Medical Records Based on Multisemantic Features by Using Robustly Optimized Bidirectional Encoder Representation From Transformers Pretraining Approach Whole Word Masking and Convolutional Neural Networks: Model Development and Validation

Weijie Wang, MS; Xiaoying Li, PhD; Huiling Ren, MS; Dongping Gao, PhD; An Fang, MS

Institute of Medical Information and Library, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China

Corresponding Author:
Huiling Ren, MS
Institute of Medical Information and Library
Chinese Academy of Medical Sciences & Peking Union Medical College
69 Dongdan N St
Beijing, 100005
China
Phone: 86 010 52328911
Email: ren.huiling@imicams.ac.cn

Abstract

Background: Clinical electronic medical records (EMRs) contain important information on patients' anatomy, symptoms, examinations, diagnoses, and medications. Large-scale mining of rich medical information from EMRs will provide notable reference value for medical research. With the complexity of Chinese grammar and blurred boundaries of Chinese words, Chinese clinical named entity recognition (CNER) remains a notable challenge. Follow-up tasks such as medical entity structuring, medical entity standardization, medical entity relationship extraction, and medical knowledge graph construction largely depend on medical named entity recognition effects. A promising CNER result would provide reliable support for building domain knowledge graphs, knowledge bases, and knowledge retrieval systems. Furthermore, it would provide research ideas for scientists and medical decision-making references for doctors and even guide patients on disease and health management. Therefore, obtaining excellent CNER results is essential.

Objective: We aimed to propose a Chinese CNER method to learn semantics-enriched representations for comprehensively enhancing machines to understand deep semantic information of EMRs by using multisemantic features, which makes medical information more readable and understandable.

Methods: First, we used Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking (RoBERTa-wwm) with dynamic fusion and Chinese character features, including 5-stroke code, Zheng code, phonological code, and stroke code, extracted by 1-dimensional convolutional neural networks (CNNs) to obtain fine-grained semantic features of Chinese characters. Subsequently, we converted Chinese characters into square images to obtain Chinese character image features from another modality by using a 2-dimensional CNN. Finally, we input multisemantic features into Bidirectional Long Short-Term Memory with Conditional Random Fields to achieve Chinese CNER. The effectiveness of our model was compared with that of the baseline and existing research models, and the features involved in the model were ablated and analyzed to verify the model's effectiveness.

Results: We collected 1379 Yidu-S4K EMRs containing 23,655 entities in 6 categories and 2007 self-annotated EMRs containing 118,643 entities in 7 categories. The experiments showed that our model outperformed the comparison experiments, with F_1-scores of 89.28% and 84.61% on the Yidu-S4K and self-annotated data sets, respectively. The results of the ablation analysis demonstrated that each feature and method we used could improve the entity recognition ability.

Conclusions: Our proposed CNER method would mine the richer deep semantic information in EMRs by multisemantic embedding using RoBERTa-wwm and CNNs, enhancing the semantic recognition of characters at different granularity levels and improving the generalization capability of the method by achieving information complementarity among different semantic features, thus making the machine semantically understand EMRs and improving the CNER task accuracy.

(JMIR Med Inform 2023;11:e44597) doi: 10.2196/44597

KEYWORDS

Chinese clinical named entity recognition; multisemantic features; image feature; Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking; RoBERTa-wwm; convolutional neural network; CNN

Introduction

Background

Abundant medical data have been accumulated since the development of the hospital information system, among which the electronic medical records (EMRs) contain information closely related to patients' diagnosis and treatment processes [1]. As important records of patients' medical activities, effective extraction and use of the medical information contained in EMRs could provide clinical decision-making support for doctors and realize personalized medical guidance and health management for patients. It could also help biomedical researchers discover the tacit medical knowledge, thus providing ideas for studies of the association between diseases, the relationship between symptoms, the prediction of diseases and therapies, complication prediction, comorbidity analysis, etc. The medical information would be rapidly extracted from the unstructured EMRs through named entity recognition (NER). NER is a basic task of natural language processing, which will lay the foundation for the construction of medical knowledge graphs, medical knowledge bases, and so on by steps such as medical entity structuring, medical entity standardization, and medical entity relationship extraction. It will also provide fundamental support for practical application scenarios such as medical knowledge retrieval systems, clinical decision support systems, clinical event extraction, and so on [2,3].

Clinical NER (CNER) refers to the recognition of entities such as anatomy, disease, symptoms, clinical examination, medication, surgical procedure, and so on from EMRs [4,5]. Chinese CNER is more difficult than English NER for several reasons. First, Chinese words lack space segmentation and have blurred boundaries. Second, the composition of a Chinese entity is complex and may contain various figures, letters, and abbreviations. Third, Chinese grammar is complicated, and the same word may represent different entity types in different contexts. Therefore, Chinese CNER remains a research focus.

Recently, the features of radicals for Chinese characters have been widely used to improve the efficiency of different Chinese natural language processing tasks [6-8]. Chinese characters, known for thousands of years, are highly developed morpheme scripts that are still used worldwide with unique ideology [9]. Chinese characters include single-component and multiple-component characters. A single-component character cannot be divided, for example, "心 (heart)," "手 (hand)," and "口 (mouth)," and so on; whereas a multiple-component character is composed of basic components, accounting for >90% of Chinese characters [10], for example, the radical for "呕 (vomit)" and "吐 (vomit)" is "口 (mouth)," and the radical for "肿 (swelling)" and "胀 (swelling)" is "月 (month)," which refers to meat or organs in ancient times. Chinese characters are divided into associative compound characters, indicative characters, pictographic characters, and picto-phonetic characters based on their characteristics. In addition, Chinese characters are also called square characters, as they are square, and there are 8 structures of Chinese characters that are subdivided based on their intrinsic shape and construction. Therefore, Chinese characters contain rich deep semantic information. Applying radicals, phonological codes, shape structures, and other features would help to improve Chinese CNER accuracy.

The contributions of this study are as follows: (1) using pretrained language model (PLM) Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking (RoBERTa-wwm) with a dynamic fusion transformer layer to obtain the semantic features of Chinese characters; (2) using CNNs for extracting the radicals and picto-phonetic features of Chinese characters through the 5-stroke code, Zheng code, phonological code, and stroke code; (3) converting Chinese characters into square images, extracting Chinese character image features from another modality by CNNs, and deeply capturing the pictographic characteristics of Chinese characters; and (4) improving the semantic recognition ability of the model at different levels of granularity, achieving information complementarity between different semantic features, and improving the effect and generalization ability of the model based on multisemantic features.

Related Works

Medical NER

In recent decades, the medical NER is still a research focus. Medical NER research has 3 main development stages as follows: based on dictionaries and rules, based on statistical machine learning, and based on deep learning.

The dictionary-based [11-13] methods need to construct a domain dictionary in advance to achieve medical NER by matching algorithms. The accuracy of this method is relatively higher. However, it may be affected by the large number, strong specialization, and high complexity of Chinese medical entities. In addition, medical terminologies are updated quickly with the rapid development of the medical field, and the lack of new terminologies will also affect medical NER accuracy. The rule-based [14,15] methods need experts in a particular field to formulate the rule templates based on information such as context grammar and structure. However, the rules are poorly universal in different fields. The methods based on dictionaries and rules [16-18] are poorly generalized, time-consuming, and objective, as much time and labor are required. Therefore, many scholars have gradually applied methods based on statistical machine learning on medical NER. The commonly used methods include maximum entropy [19], support vector machine [20,21], hidden Markov model [22,23], and conditional random fields (CRF) [24,25]. However, these methods rely on large-scale annotation data sets [26] and manual feature selection [13,27]. Moreover, the quality of the selected features will directly affect the medical NER results.

With the continuous development of deep learning, Cocos et al [28] found that deep learning has advantages over traditional machine learning. It can automatically extract the characteristics of various levels and reduce the subjectivity of artificial feature selection. This thereby improves the result accuracy. The commonly used deep learning models include convolutional neural networks (CNNs) [29], recurrent neural networks [30], long short-term memory (LSTM) [31], Word to Vector (Word2Vec) [32], Bidirectional Encoder Representation from Transformers (BERT) [33], and so on. However, fully extracting the data features by using a neural network alone is challenging. Most scholars took Long LSTM-CRF as the main framework to make up for the medical NER deficiency using a single neural network [34]. The Bidirectional LSTM-CRF (BiLSTM-CRF) [35] model was then developed. This model could better capture contextual information as an important milestone in medical NER and has been widely used in the medical field [36,37]. To improve the ability to capture details and extract features of medical NER models, many studies added Word2Vec with static representation [38], Global Vectors for Word Representation [39] with static representation, Embeddings from Language Models (ELMo) [40,41] with dynamic representation, CNN [42], and attention mechanism [43] to the BiLSTM-CRF model. Some studies [44,45] have shown that the application of the BiLSTM-CRF model combined with the word vector generated by BERT could significantly improve medical NER accuracy. BERT provided a more accurate word representation and achieved better task results than traditional word vector methods. As per the specialty of medicine and the characteristics of Chinese characters involved, the clinical dictionaries, root-level features, parts of speech, radicals, and phonological codes have been added in the BiLSTM-CRF model in some studies [46-51] for improving Chinese CNER performance.

PLMs Technique

PLMs are pretrained on a large-scale corpus to obtain prior semantic knowledge from unlabeled text and improve the effectiveness of different downstream tasks. The word vector generated by a bidirectional language model BERT with stacked transformer substructures contains not only the preliminary information from the corpus training but also the encoded contextual information. Some robust versions of BERT have been constructed since BERT was proposed in 2018. For example, the RoBERTa model [52], which replaces the static (MASK) strategy with a dynamic (MASK) strategy, and the words (MASK) in each sequence dynamically change in different epoch trainings. In addition, the RoBERTa model is retrained with bigger batches and longer sequences, and the next-sentence prediction task, which is not related to the downstream task, is canceled during the pretraining. Compared with the BERT model, the RoBERTa model performs better on multiple natural language processing tasks. However, the character-level RoBERTa model does not fit the Chinese natural language processing, as the different segmentation modes between Chinese and English words suffer a limitation of lacking word information. Then, the word-level RoBERTa-wwm model [53] was proposed based on Chinese characteristics, which greatly improved the text representation ability in Chinese [54].

Methods

Data Collection

The Yidu-S4K data set, shared publicly by YiduCloud, is derived from the Chinese EMRs entity recognition task of the China Conference on Knowledge Graph and Semantic Computing 2019 [55]. It contains 1379 EMRs with 6 entity types, including Disease (medically defined disease and diagnoses made by physicians based on etiology, pathophysiology, pathological classification, and clinical staging); Anatomy (anatomical parts of the body where disease, signs, and symptoms occurred); Laboratory (physical or chemical tests performed by the laboratory department in clinical work); Image (imaging [x-ray, computed tomography, magnetic resonance imaging, positron emission tomography-computed tomography, etc], ultrasound, and electrocardiogram); Medicine (specific chemical substances used for disease treatment); and Operation (treatments focused on surgery such as excision and suturing performed by the physician locally on the patient’s body).

Self-annotated EMR data, collected from publicly desensitized Chinese EMR websites [56], contain 2007 EMRs. As per the Terminology of Clinical Medicine issued by the National Health Commission of the People’s Republic of China, we used the BIO (B signifies the beginning of an entity, I signifies that the word is inside an entity, and O signifies that the word is just a regular word outside of an entity) tagging method to pretag 7 entity types in the EMRs, including Disease (same definition as the Yidu-S4K data set); Symptoms (abnormal manifestations as perceived by the sensory organs of patients and physicians); Anatomy (same definition as the Yidu-S4K data set); Examination (includes imaging examinations and laboratory tests mentioned in the Yidu-S4K data set); Instrument (apparatus and mechanical equipment for disease prevention, diagnosis, treatment, health care, and rehabilitation); Medicine (the same definition as the Yidu-S4K data set); and Operation (same definition as the Yidu-S4K data set). Subsequently, 4 medical experts manually checked and corrected the tags. The interclass correlation efficient consistency test revealed that we had good annotation quality.

The ratio of the training set to the test set of the EMRs was 7:3. The Yidu-S4K data set was preprovided with 1000 EMRs as the training data sets (1000/1379, 72.52%) and 379 EMRs as the test data sets (379/1379, 27.48%). The self-annotated data set was divided by randomization into 1401 EMRs as the training data sets (1401/2007, 69.81%) and 379 EMRs as the test data sets (606/2007, 30.19%). Table 1 lists the details of the different types of entities in the 2 EMR data sets.

Table 1. The statistics of different types of entities in 2 electronic medical record data sets

Data sets and entity type	Training set, n	Test set, n
Yidu-S4K		
Disease	4212	1323
Anatomy	8426	3094
Laboratory	1195	590
Image	969	348
Medicine	1822	485
Operation	1029	162
All entities	17,653	6002
Self-annotated		
Disease	9470	4504
Symptoms	26,334	11,065
Anatomy	17,877	7588
Examination	19,664	8746
Instrument	1244	560
Medicine	5314	2566
Operation	2578	1133
All entities	82,481	36,162

Ethical Considerations

Ethics approval was not required because the patient's private information was masked by the website.

Experiments Settings

In this study, all the experiments were conducted by Python [57] and PyTorch [58]. Table 2 shows the experimental parameters. The experiments used *RoBERTa-wwm-ext-large* model pretraining data, optimized parameters using Adam W, dropout to prevent overfitting, the batch size of 32, BiLSTM hidden layer dimension of 768, maximum sequence length of 510, RoBERTa-wwm dimension of 768, semantic feature dimension of 124, and image feature dimension of 128. On 2 Chinese CNER data sets, we used the same parameters.

Table 2. Parameter settings.

Parameter	Value
Dropout	0.5
Epoch	Optimization
Optimization	Adam W
Learning rate	0.0001
Batch size	32
BiLSTM[a] hidden layer	768
Max_len	510
RoBERTa-wwm[b] feature dimension	768
Semantic feature dimension	124
Image feature dimension	128

[a] BiLSTM: Bidirectional Long Short-Term Memory.

[b] RoBERTa-wwm: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking.

Evaluation Metrics

The experiments used precision, recall, and F_1-score to evaluate the model performance. The formulas for each index are as follows:

Precision = TP / (TP + FP) **(1)**

Recall = TP / (TP + FN) **(2)**

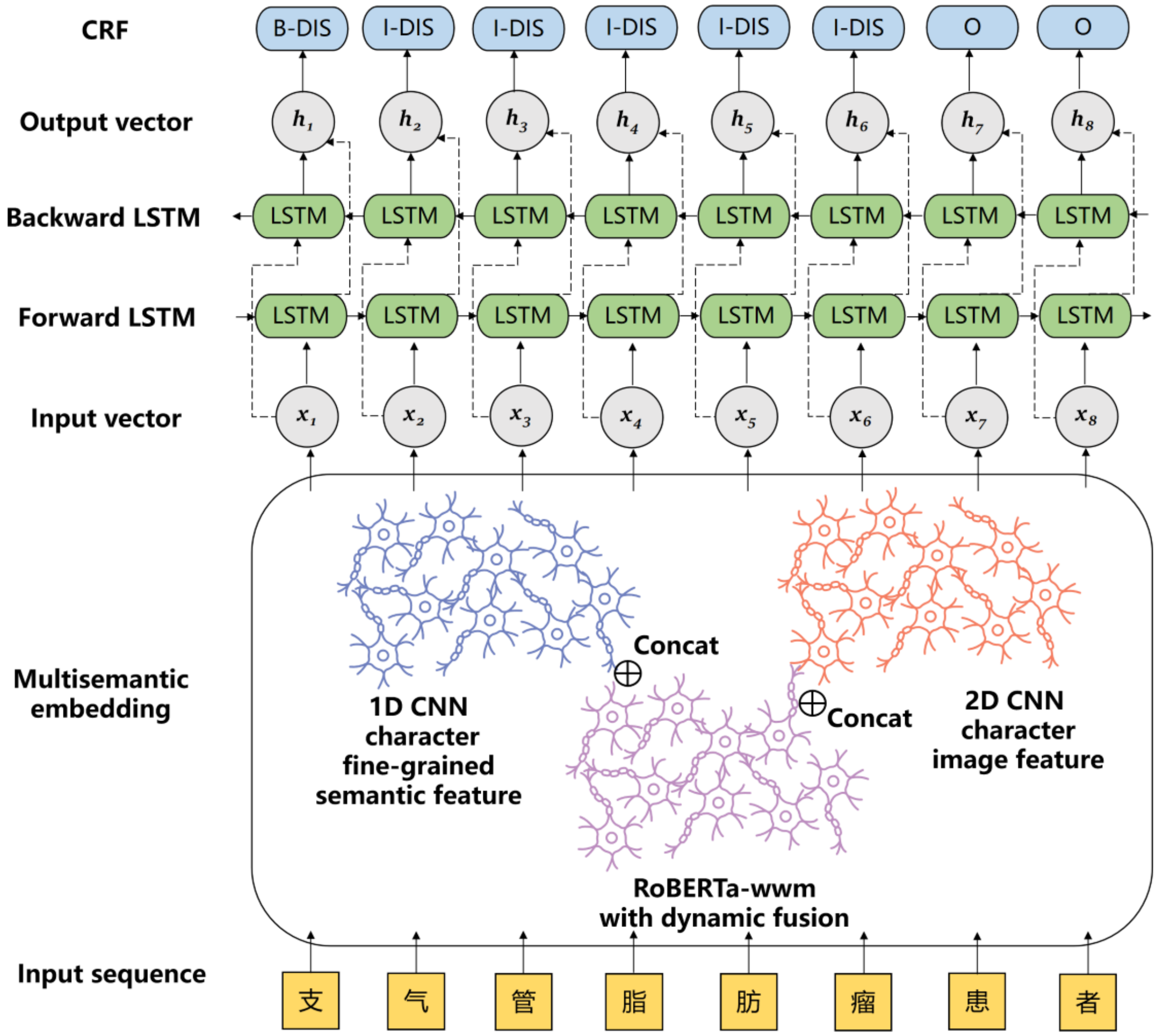
F_1-score = (2 × precision × recall) / (precision + recall) **(3)**

where precision is the proportion of positive samples in all samples predicted to be positive; recall is the proportion of positive samples in all positive samples; F_1-score is the harmonic mean of precision and recall; true positive (TP) is the number of positive samples predicted to be positive, that is, the number of correctly recognized entities; false positive (FP) is the number of negative samples predicted to be negative, that is, the number of incorrectly recognized other texts as entities; and false negative (FN) is the number of positive samples predicted to be negative, that is, the number of unrecognized entities.

Model Overview

In this study, we proposed a CNER model based on multisemantic features, as shown in Figure 1. First, we used RoBERTa-wwm, the PLM, to obtain the embedded representation at the word level. Dynamic fusion is performed on the semantic representation generated by each transformer layer to make full use of RoBERTa-wwm representation information. Then, the embedded Chinese character fine-grained feature representation, including the 5-stroke code, Zheng code, phonological code, and stroke code, is extracted by 1D CNN, whereas the embedded Chinese character image representation is extracted from another modality by 2D CNN, with the Chinese characters as square images. Finally, the above multisemantic vectors were input into the BiLSTM layer for encoding and were decoded in the CRF layer to predict the tag probability.

Figure 1. The main architecture of our model. 1D CNN: 1D convolutional neural network; 2D CNN: 2D convolutional neural network; B-DIS: beginning of disease entity; CRF: conditional random fields; h: embedding of output character; I-DIS: inside of disease entity; LSTM: long short-term memory; O: other type; RoBERTa-wwm: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking; x: embedding of input character.



Multisemantic Embedding Layer

Overview

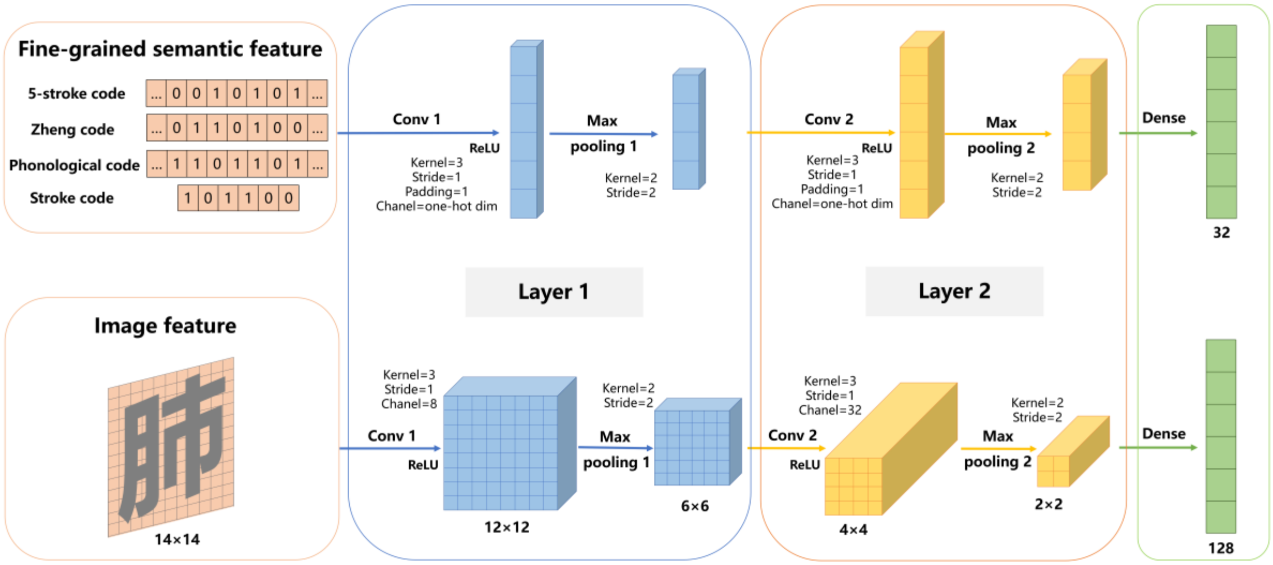
Many Chinese characters have retained their original connotations, as they originated from pictographic characters in ancient times. Moreover, the inherent fine-grained character information contained in Chinese characters often implies more additional semantic information. Accordingly, we obtained the 5-stroke code, Zheng code, phonological code, and stroke code information, as shown in Table 3, of the Chinese characters from ZDIC [59] and embedded them in the model. In addition, Chinese characters are squares, and different shapes and structures express different types of information. Characters with similar intrinsic characteristics may have similar meanings. Therefore, we took Chinese characters as graphics and obtained semantic information on Chinese character connotations from another modality. Multisemantics could obtain information comprehensively and learn a better feature representation by making use of information complementarity and eliminating the redundancy among different semantic features compared with a single-semantic feature, resulting in a more generalized model.

As shown in Figure 2, we converted the Chinese character 5-stroke code, Zheng code, phonological code, and stroke code into one-hot vector encoding and interpreted the Chinese characters as 14×14 images. Subsequently, we used a 2-layer CNN deeply extracting the Chinese character multisemantic features. Through the Convolution layer with the *ReLU* activation function, max pooling layer, and dense layer, we obtained the multisemantic vectors that could be embedded in the BiLSTM layer.

Table 3. Example of Chinese characters' coded information from ZDIC.

Character	5-stroke code	Zheng code	Phonological code	Stroke code
呕 (vomit)	kaqy	jhos	ǒu	2,511,345
吐 (vomit)	kfg	jbvv	tù	251,121
肿 (swelling)	ekhh	qji	zhǒng	35,112,512
胀 (swelling)	etay	qch	zhàng	35,113,154
心 (heart)	nyny	wz	xīn	4544
手 (hand)	rtgh	md	shǒu	3112

Figure 2. The process of obtaining Chinese character multisemantic features by convolutional neural network. ReLU: Rectified Linear Unit function; Conv 1: first convolutional layer; Conv 2: second convolutional layer; Max pooling 1: first max pooling layer; Max pooling 2: second max pooling layer; Dense: dense layer.



RoBERTa-wwm With Dynamic Fusion

When RoBERTa-wwm pretrains the corpus, it is segmented on the language technology platform established by the Harbin Institute of Technology based on Wikipedia content in Chinese, which can provide a basis for achieving wwm. As shown in Figure 3, the word "支气管 (bronchi)" in the RoBERTa-wwm model is completely masked by random wwm, whereas only single characters can be randomly masked in the BERT model, for example, only 1 character "气 (gases)" was masked in the word "支气管 (bronchi)." Thus, the RoBERTa-wwm model can learn the word-level semantic representations in Chinese.

The encoder structure of each transformer layer of the BERT model outputs had different abstract representations of grammar, semantics, and real knowledge in sentences. Studies have confirmed that each layer of the BERT model represents text information differently through 12 natural language processing tasks [60]. As shown in Figure 4, the low transformer mainly learns and encodes surface features; the middle transformer learns and encodes syntactic features; and the high transformer learns and encodes semantic features.

The transformer structure of the RoBERTa-wwm model is consistent with that of the BERT model. To make full use of the representation information of each transformer layer, we used the RoBERTa-wwm model with dynamic fusion [61]. This helped in assigning the initial weight to the representation vector of 12 transformer layers, determining the weight during training, and weighing the representation vector generated by each layer.

Figure 3. Mask process of Bidirectional Encoder Representation from Transformers (BERT) and Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking (RoBERTa-wwm).

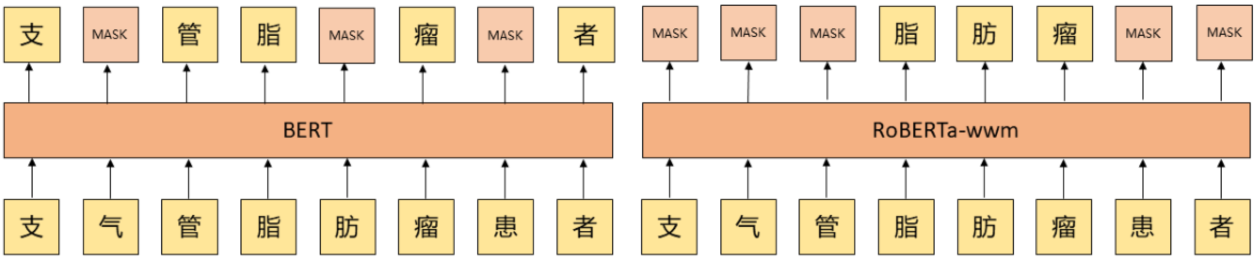
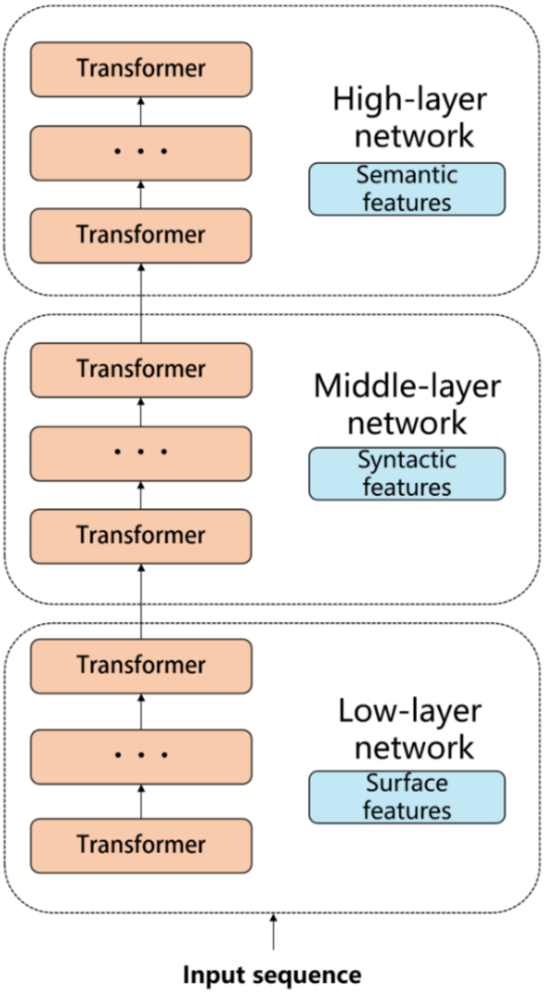


Figure 4. Coding representation of Transformer with 12 layers of Bidirectional Encoder Representation from Transformers model.



Assume that the text input sequence *seq* = (x_1,x_2,x_3, …, x_n), where n is the total length of the character contained in the sequence; x_i is the i^{th} character of the input sequence; and the fusion formula is as follows:

$$v_i^{RoBERTa\text{-}wwm}=Dense_{unit=512}\ (x_i)\ \{\textstyle\sum^{n}_{i=1}\alpha_c \times h_c\},\ (c \in [1,12])\ \textbf{(4)}$$

$v_i^{(RoBERTa\text{-}wwm)}$ is the output representation by the RoBERTa-wwm model with dynamic fusion for the current character x_i; h_c is the output representation by each transformer layer of the RoBERTa-wwm model, and α_c is the output weight value assigned to each layer by RoBERTa-wwm.

Fine-Grained Semantic Feature

5-Stroke Code

The 5-stroke code is a typical semantic code, which encodes Chinese characters according to strokes and structures. Currently, it is widely used to code Chinese characters. The expression of the 5-stroke code may inevitably repeat with the phonological code, for example, the 5-stroke code for “亦 (also)” is “you,” while the phonological code for “亦 (also)” is also “you” [62]. Hence, we combined the 5-stroke code and Zheng code to compensate for the encoding deficiency. We used the 5-stroke code in Zdic.net to vectorize the Chinese characters using the following formulas:

$$p = f_{fc}\ (seq)\ \mathbf{(5)}$$

$$vifc=efc\ (pi),\ (i\in Z \cap i\in\ [1,n])\ \mathbf{(6)}$$

where f_{fc} represents the function that maps the input character sequence into the 5-stroke code and v_i^{fc} represents the 5-stroke code vector corresponding to x_i.

Zheng Code

The Zheng code was created by famous Chinese literature professors as per the strokes and roots of Chinese characters through in-depth research on the patterns and structures of Chinese characters. The early Microsoft operating system in Chinese adopted the Zheng code as the built-in code. This indicates that Zheng code is a scientific coding of Chinese characters. Chinese characters with similar codes may contain related semantic information. Hence, the potential semantic relationship of text may be found by mining the structural information of Chinese characters using Zheng code. The Zheng code was vectorized as the 5-stroke code and has the following formulas:

$$p = f_{zc}\ (seq)\ \mathbf{(7)}$$

$$vizc=ezc\ (pi),\ (i\in Z \cap i\in\ [1,n])\ \mathbf{(8)}$$

where f_{zc} represents the function that maps the input character sequence into the Zheng code and v_i^{zc} represents the Zheng code vector corresponding to x_i.

Phonological Code

Over 90% of Chinese characters are picto-phonetic characters [63]. Hence, pronunciation plays an important role in the semantic expressions of Chinese characters. We used the Pypinyin toolkit to vectorize the phonological code of Chinese characters using the following formulas:

$$p = f_{pc}\ (seq)\ \mathbf{(9)}$$

$$vipc=epc\ (pi),\ (i\in Z \cap i\in\ [1,n])\ \mathbf{(10)}$$

where f_{pc} represents the function that maps the input character sequence into the phonological code and v_i^{pc} represents the phonological vector corresponding to x_i.

Stroke Code

Chinese characters with similar strokes may have similar meanings. The strokes of each Chinese character were encoded in ZDIC [59], where 1, 2, 3, 4, and 5 represent the horizontal stroke, vertical stroke, left-falling stroke, right-falling stroke, and turning stroke, respectively. We transformed the stroke code into a 5-dimension vector, where each dimension was the corresponding number of strokes. The stroke code was vectorized in the same manner as the 5-stroke code and has the following formulas:

$$p = f_{sc}\ (seq)\ \mathbf{(11)}$$

$$visc=esc\ (pi),\ (i\in Z \cap i\in\ [1,n])\ \mathbf{(12)}$$

where f_{sc} represents the function that maps the input character sequence into the stroke code and v_i^{sc} represents the stroke code vector corresponding to x_i.

To extract the fine-grained semantic features of Chinese characters deeply, we trained the character features using CNNs. The character features were trained by 2 convolutions with a kernel of 3 and *ReLU* function as well as max pooling of 2×2, where the number of output channels was the dimension of each feature vector. Finally, the 32-dimension Chinese character vector was obtained through a full connection in the dense layer, as shown in Figure 2.

Image Feature

Chinese characters have been derived from pictographic symbols since ancient times, and characters with similar symbolic appearances have similar image features. However, the fonts of Chinese characters have changed a lot over time. Simplified characters have lost much pictographic information compared with complex characters. Therefore, Cui et al [64] used Chinese character images to extract Chinese character features and achieved better performance. Wu et al [65] tried different character fonts and found that the best result was obtained by using the *NotoSansCJKsc-Regular* font. On the basis of these findings, we used the Python Imaging Library to convert *NotoSansCJKsc-Regular* Chinese characters into black-and-white images and extracted image features by 2D CNN in depth as per the following formulas:

$$e^{if\_1}=(Max\ pooling\ 1\ (Conv\ 1\ (K \otimes H))\ \mathbf{(13)}$$

$$e^{if\_2}=(Max\ pooling\ 2\ (Conv\ 2\ (K \otimes H)\ \mathbf{(14)}$$

$$v_i^{if}=\text{Dense}\ (e^{if\_2})\ \mathbf{(15)}$$

where K is a kernel; H is the original embedded image matrix; *Conv 1*, *Max pooling 1*, *Conv 2*, and *Max pooling 2* are the first convolutions with a kernel of 3 and channel of 8, the first max pooling with the kernel of 2×2, the second convolution with the kernel of 3 and channel of 32, and the second max pooling with the kernel of 2×2, respectively; e^{if_1} is the result after the first convolution; e^{if_2} is the result after the second convolution; Dense is the process of realizing the full connection; and v_i^{if} is the final 128-dimension Chinese character image vector trained by convolution, as shown in Figure 2.

Finally, the multisemantic features $v_i^{RoBERTa\text{-}wwm}$, v_i^{fc}, v_i^{zc}, v_i^{pc}, v_i^{sc}, and v_i^{if} were embedded by the array Concat function. The formula used is as follows:

$$v_i^{input}=\text{Concat}\ (v_i^{RoBERTa\text{-}wwm},\ v_i^{fc},\ v_i^{zc},\ v_i^{pc},\ v_i^{sc},\ v_i^{if})\ \mathbf{(16)}$$

BiLSTM Layer

The role of BiLSTM [66] is essential in NER. As shown in Figure 1, the forward LSTM and backward LSTM are responsible for memorizing the previous and subsequent text information, respectively. By combining the 2, contextual information can be obtained simultaneously, which helps to capture the bidirectional semantic dependency information in the text. The formulas used are as follows:

$$h_i^{forward}=LSTM^{forward}(\alpha^{<i-1>}, x_i) \quad \mathbf{(17)}$$

$$h_i^{backward}=LSTM^{backward}(\alpha^{<i-1>}, x_i) \quad \mathbf{(18)}$$

$$h_i=[h_i^{forward}; h_i^{backward}] \quad \mathbf{(19)}$$

where $\alpha^{<i>}$ represents the hidden layer state of the current memory cell; $LSTM^{forward}$ is the feature representation from front to back; $LSTM^{backward}$ is the feature representation from back to front; $h_i^{forward}$ is the forward semantic information obtained through the forward LSTM at the i-th character position; $h_i^{backward}$ is the backward semantic information obtained through the backward LSTM at the i-th character position; and h_i represents a combination of hidden states in both.

CRF Layer

The BiLSTM can be used to handle contextual relationships. However, it cannot consider the dependencies between tags. Therefore, it is necessary to add a constraint relation for the final predicted label by using the CRF [67] layer to ensure the predicted label rationality. Given an input sequence where $X=\{x_1,x_2,...,x_n\}$, we assume that the training output label sequence is $Y=\{y_1,y_2,...,y_n\}$, where n is the number of model labels. The sequence score of the label and the probability of the label sequence y are calculated as follows:

$$P\ (y|X)=e^{\Sigma i=1}\ (Z_{yi,yi+1}\ +\ P_{i+1,yi+1})\ /\ (\Sigma_{(y\ \tilde{}\in Yx)}\ e^{\Sigma i=1}\ (Z_{yi,yi+1}\ +\ P_{i+1,yi+1}) \quad \mathbf{(20)}$$

where Z is the transfer matrix; $Z_{yi,yi+1}$ is the score of the label transfer from y_i to y_{i+1}; $P_{i+1,yi+1}$ is the score of label y_{i+1} corresponding to the $i+1$th character of the input sequence; Y_x is the set of all possible label sequences. The final label of the output sequence is the set of labels with the highest probability.

Finally, we predicted the best label sequences by using the Viterbi algorithm [68] with the following formula:

$$y^*=argmax(s(X, y)) \quad \mathbf{(21)}$$

Results

To get convincing experimental results, we ran each model 5 times and calculated the average precision, average recall, and average F_1-score.

Performance Comparison With Ensemble Models

To verify the validity of the model, we compared our model with the existing ensemble models BiLSTM-CRF, ELMo-Lattice-LSTM-CRF, ELMo-BiLSTM-CRF, all CNNs, ELMo-encoder from transformer-CRF, and multigranularity semantic dictionary and multimodal tree-NER on Yidu-S4K and self-annotated data sets, and the results are shown in Table 4. The F_1-scores of the experimental model on the Yidu-S4K data set were 18.31%, 4.15%, 4.26%, 4.12%, 3.69%, and 2.59% higher than those of the BiLSTM-CRF, all CNNs, ELMo-Lattice-LSTM-CRF, ELMo-BiLSTM-CRF, ELMo-encoder from transformer-CRF, and multigranularity semantic dictionary and multimodal tree-NER models, respectively. On the self-annotated data set, it was 5.14% higher than that of the BiLSTM-CRF. The results showed that the performance of the experimental model is superior to that of the existing model.

Table 4. Performance comparison of ensemble models on the Yidu-S4K and self-annotated data sets.

Data set and model	Precision (%)	Recall (%)	F_1-score (%)
Yidu-S4K			
BiLSTM-CRF[a] [64]	69.43	72.58	70.97
ACNN[b] [69]	83.07	87.29	85.13
ELMo[c]-lattice-LSTM-CRF [70]	84.69	85.35	85.02
ELMo-BiLSTM-CRF [41]	—[d]	—	85.16
ELMo-ET[e]-CRF [71]	82.08	86.12	85.59
MSD_DT_NER[f] [72]	86.09	87.29	86.69
Our model	90.37	88.22	89.28
Self-annotated			
BiLSTM-CRF	81.98	77.10	79.47
Our model	84.24	84.99	84.61

[a] BiLSTM-CRF: Bidirectional Long Short-Term Memory-conditional random fields.

[b] ACNN: all convolutional neural network.

[c] ELMo: Embeddings from Language Models.

[d] Not available.

[e] ET: encoder from transformer.

[f] MSD_DT_NER: multigranularity semantic dictionary and multimodal named entity recognition.

Performance Comparison With PLMs Related to BERT

The performance of the PLM, BERT, is a milestone in natural language processing. To verify the BERT robust version’s validity of the RoBERTa-wwm model, we compared our model with the existing ensemble models with the BiLSTM-CRF, BERT-BiLSTM-CRF, and RoBERTa-wwm-BiLSTM-CRF on Yidu-S4K and self-annotated data sets, and the results are shown in Table 5. The F_1-scores of the experimental model on the Yidu-S4K data set were 18.31%, 2.99%, and 0.82% higher than those of the BiLSTM-CRF, BERT-BiLSTM-CRF, and RoBERTa-wwm-BiLSTM-CRF models, respectively, and 5.14%, 2.95%, and 1.07% higher on the self-annotated data set, respectively.

Table 5. Performance comparison of PLMs[a] on the Yidu-S4K and self-annotated data sets.

Data set and model	Precision (%)	Recall (%)	F_1-score (%)
Yidu-S4K			
BiLSTM[b]-CRF[c] [64]	69.43	72.58	70.97
BERT[d]-BiLSTM-CRF	89.07	83.67	86.29
RoBERTa-wwm[e]-BiLSTM-CRF	90.08	86.90	88.46
Our model	90.37	88.22	89.28
Self-annotated			
BiLSTM-CRF	81.98	77.10	79.47
BERT-BiLSTM-CRF	82.48	80.86	81.66
RoBERTa-wwm-BiLSTM-CRF	84.23	82.86	83.54
Our model	84.24	84.99	84.61

[a] PLM: pretrained language model.

[b] BiLSTM: Bidirectional Long Short-Term Memory.

[c] CRF: conditional random fields.

[d] BERT: Bidirectional Encoder Representation from Transformers.

[e] RoBERTa-wwm: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking.

Performance Comparison of Each Entity

To comprehensively evaluate our model, we calculated the F_1-score for each entity type on the Yidu-S4K and self-annotated data sets, as shown in Tables 6 and 7. The F_1-score of our model on the Yidu-S4K data set for each of the 6 entity categories, except for the Image entity, increased by 0.2% to 7.6% compared with the data listed in the tables. The F_1-score for the Image entity was 0.35% lower than that of the ELMo-BiLSTM-CRF model. However, the F_1-scores for the Laboratory entity and Operation entity were 7.6% and 7.54% higher than those of the ELMo-BiLSTM-CRF model, respectively. The overall F_1-score was 4.12% higher than that of the ELMo-BiLSTM-CRF model. For the self-annotated data set, our model improved each entity in 7 categories ranging from 0.09% to 14.49% over the listed data, with a greater improvement for Instrument entities.

Table 6. Performance comparison of each entity category on the Yidu-S4K data set.

Model	F_1-score for each category (%)						
	All	Disease	Anatomy	Image	Laboratory	Medicine	Operation
ELMo[a]-BiLSTM[b]-CRF[c] [41]	85.16	82.81	85.99	88.01	75.65	94.49	86.79
BERT[d]-BiLSTM-CRF	86.29	87.14	86.36	83.43	77.98	89.46	93.11
BERT-wwm[e]-BiLSTM-CRF	87.12	86.18	85.47	81.52	79.69	90.14	92.49
RoBERTa[f]-wwm-BiLSTM-CRF	88.46	87.71	87.01	86.69	82.36	93.22	92.87
Our model	89.28	87.91	87.47	87.66	83.25	94.98	94.33

[a] ELMo: Embeddings from Language Models.

[b] BiLSTM: Bidirectional Long Short-Term Memory.

[c] CRF: conditional random fields.

[d] BERT: Bidirectional Encoder Representation from Transformers.

[e] wwm: Whole Word Masking.

[f] RoBERTa: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach.

Table 7. Performance comparison of each entity category on the self-annotated data set.

Model	F_1-score for each category (%)							
	All	Disease	Symptoms	Anatomy	Examination	Instrument	Medicine	Operation
BERT[a]-BiLSTM[b]-CRF[c]	81.66	81.33	85.87	83.86	90.36	60.38	89.72	79.75
BERT-wwm[d]-BiLSTM-CRF	81.58	74.91	83.89	81.23	88.84	54.76	85.63	68.49
RoBERTa[e]-wwm-BiLSTM-CRF	83.54	81.99	86.69	84.68	91.21	66.01	91.04	81.17
Our model	84.61	82.34	86.93	85.62	91.30	69.25	91.28	82.49

[a] BERT: Bidirectional Encoder Representation from Transformers.

[b] BiLSTM: Bidirectional Long Short-Term Memory.

[c] CRF: conditional random fields.

[d] wwm: Whole Word Masking.

[e] RoBERTa: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach.

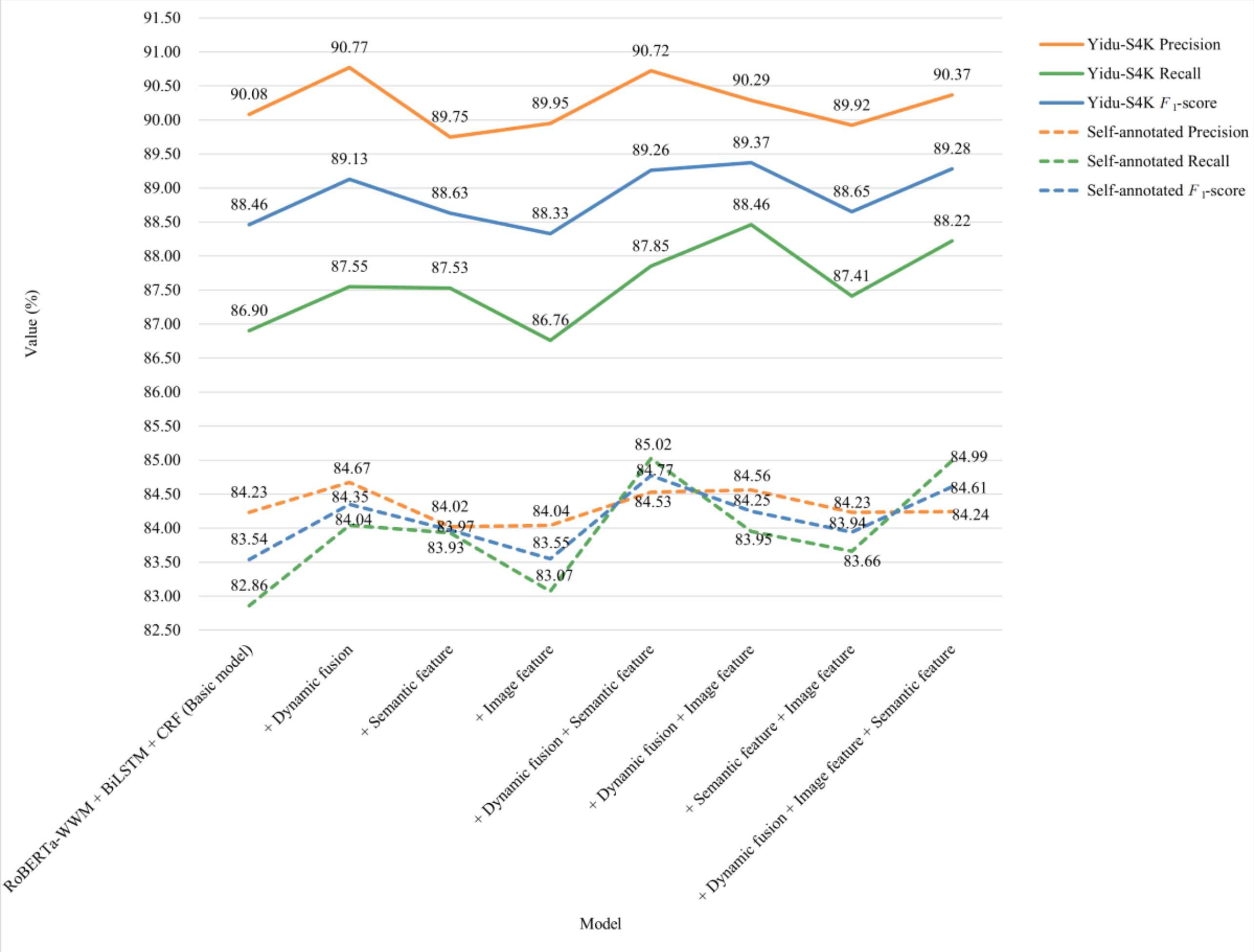
Ablation Analysis

Ablation Experiments for Multisemantic Features

To verify the fine-grained semantic features and image features of Chinese characters, dynamic fusion was effective. We used the RoBERTa-wwm-BiLSTM-CRF model as the baseline to perform ablation experiments for the above contents on 2 EMR data sets, and the results are shown in Figure 5.

The performance of the model was significantly improved with the dynamic fusion of RoBERTa-wwm. After incorporating the semantic features of Chinese characters into the model alone, the overall performance of the model was not as high as that after dynamic fusion. However, the performance on both data sets was superior to that of the baseline. The performance of the model was unstable when image features of Chinese characters were added to the model alone. On the Yidu-S4K data set, the model's performance was inferior to that of the baseline, whereas on the self-annotated data set, the model's performance only improved slightly. After adding the semantic and image features of Chinese characters to the model, the performance of the model on the Yidu-S4K data set was superior to that of the baseline. Furthermore, it was better than that of the model with semantic or image features of Chinese characters alone. The performance of the model on the self-annotated data set was superior to that of the baseline and better than that of the model with the image features of Chinese characters alone. When the model combined dynamic fusion with the semantic features and image features of Chinese characters, it was found that the performance of the model was significantly improved on the 2 data sets. Dynamic fusion with image features of Chinese characters showed the best comprehensive performance on the Yidu-S4K data set, whereas dynamic fusion with semantic features of Chinese characters achieved the best comprehensive performance on the self-annotated data set. After combining the semantic and image features of the Chinese characters and dynamic fusion, it was noted that the performance of the model was superior to that of the baseline. Because the quality of the self-annotated EMRs is inferior to that of the public Chinese EMRs corpus and the self-annotated data set contains a wider coverage of departments, the comprehensive effect of the self-annotated data set is lower than that of the YiduS4K data set in Figure 5.

Figure 5. The results of ablation experiments for mutisemantic features on the Yidu-S4K and self-annotated data sets. BiLSTM: Bidirectional Long Short-Term Memory; CRF: Conditional Random Fields; RoBERTa-wwm: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking.



Ablation Experiments for Fine-Grained Semantic Features

The fine-grained semantic features of Chinese characters used in this study included the 5-stroke code, Zheng code, phonological code, and stroke code. To verify the effectiveness of these features, we used the RoBERTa-wwm-BiLSTM-CRF model as the baseline to perform ablation experiments for the 4 features on the 2 EMR data sets, and the results are shown in Figure 6 and Figure 7.

Figure 6. The results of ablation experiments for fine-grained semantic features on the Yidu-S4K data set. BiLSTM: Bidirectional Long Short-Term Memory; CRF: conditional random fields; RoBERTa-wwm: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking.

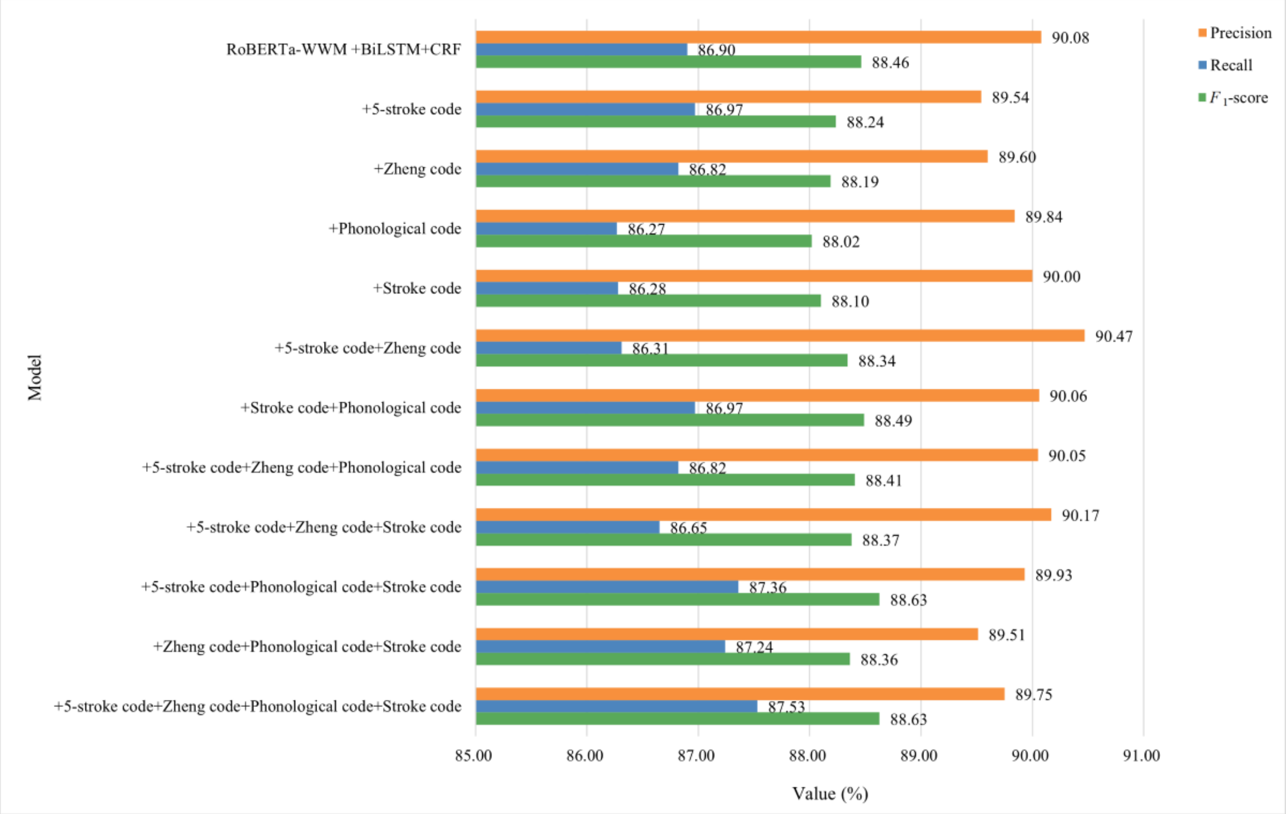
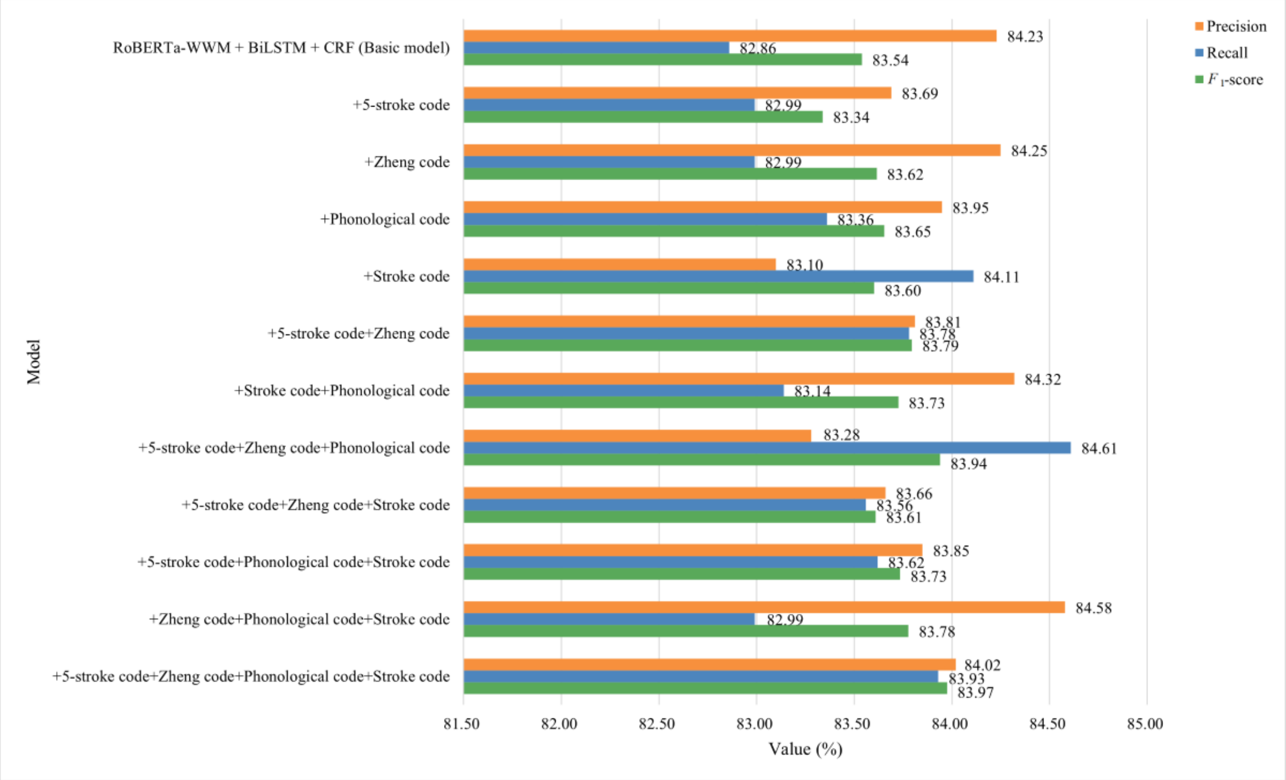


Figure 7. The results of ablation experiments for fine-grained semantic features on the self-annotated data set. BiLSTM: Bidirectional Long Short-Term Memory; CRF: conditional random fields; RoBERTa-wwm: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach Whole Word Masking.



The F_1-score of the model on the Yidu-S4K data set ranked in the top 2 for the 5-stroke code and Zheng code, whereas the F_1-score on the self-annotated data set ranked in the top 2 for the phonological code or Zheng code. The performance of the model combining 2 features (the combination of 5-stroke code and Zheng code or the combination of phonological code and stroke code) was better than that of the model with only 1 feature, regardless of the data set. On the Yidu-S4K data set, the model combining 5-stroke code+phonological code+stroke code showed the best comprehensive performance, followed by the combinations of 5-stroke code+Zheng code+phonological code, 5-stroke code+Zheng code+stroke code, and Zheng code+phonological code+stroke code. On the self-annotated data set, the model combining the 5-stroke code+Zheng code+phonological code showed the best comprehensive performance, followed by Zheng code+phonological code+stroke code, 5-stroke code+phonological code+stroke code, and 5-stroke code+Zheng code+stroke code. On the Yidu-S4K data set, only the model combining the 5-stroke code+phonological code+stroke code showed a comprehensive performance superior to that of the baseline. However, on the self-annotated data set, the comprehensive performance of all combinations was superior to that of the baseline. The performance of the model combining 3 features was less stable. The model combined 4 features on the Yidu-S4K and self-annotated data sets and achieved the best comprehensive performance among all the combinations.

Error Analysis

From Tables 6 to Table 7, our model improved the entity recognition performance of each entity category to different degrees. However, the entity recognition effect differs for each category. The F_1-scores of Disease, Anatomy, Image, Laboratory, Medicine, and Operation entity recognition on the Yidu-S4K data set were 87.91%, 87.47%, 87.66%, 83.25%, 94.98%, and 94.33%, respectively. The F_1-scores of Disease, Symptoms, Anatomy, Examination, Instrument, Medicine, and Operation entity recognition on the self-annotated data set were 82.34%, 86.93%, 85.62%, 91.31%, 69.25%, 91.28%, and 82.49%, respectively. On the Yidu-S4K data set, the precision of Laboratory entity recognition was the lowest, followed by the Anatomy entity, Image entity, and Disease entity. On the self-annotated data set, the precision of Instrument entity recognition was the lowest, followed by the Disease entity, Anatomy entity, and Operation entity. We concluded the following 7 main causes of the errors that occurred based on a review of the data set and model prediction results, as shown in Table 8.

We strictly controlled the annotation quality of both data sets. Hence, the probability of causes (1-3) was relatively low. Causes (4-6) were more likely to occur, and cause (7) mainly occurred on some entities that were less common or had fewer training samples.

Table 8. Different types of errors on 2 data sets.

Types of errors Illustrations	Example
(1) Annotation error	
1. Some manually annotated entities contained punctuation marks unrelated to the entities.	For instance, some Laboratory entities, like “PLT[a],” “NEUT[b],” and “CAE[c],” on the Yidu-S4K data set contained commas, which were correctly recognized as “PLT,” “NEUT,” and “CAE” in our model.
2. A few entity categories were confused.	For example, “PET-CT[d]” was manually annotated as a Laboratory entity on the Yidu-S4K data set, but our model correctly predicted as an Image entity.
(2) Inconsistent annotation	
The inconsistent annotation will affect the accuracy of machine learning.	On the Yidu-S4K data set, the character “下 (below)” expressing orientation of “剑突下 (below xiphoid)” was not annotated, and the character “部 (part)” expressing the part of “咽喉部 (hypopharynx)” was also not annotated. Most of the characters expressing specific locations were annotated.
(3) Missing annotation	
The missing annotated entity will also affect the overall effect of the model.	The Disease entity “窦性心律 (sinus rhythm)” was missed annotated on the Yidu-S4K data set, and the Medicine entity “生理盐水 (normal saline)” was missed annotated on the self-annotated data set.
(4) Entity with a non-Chinese character symbol	
Figures, letters, and other symbols cannot be extracted with more semantic features than Chinese characters. Hence, it may be difficult to recognize entities with symbols other than Chinese characters in the Chinese corpus.	The failure to recognize the non-Chinese character entities, like the Laboratory entity “AFP[e]” on the Yidu-S4K data set and the Examination entity “HCG[f]” on the self-annotated data set, so did such situations as the Medicine entity “VP[g]-16” was recognized as “VP-,” and “50%葡萄糖 (50% glucose)” as “葡萄糖 (glucose)” on the Yidu-S4K data set.
(5) Presence of nested entities	
On the Yidu-S4K data set, the Disease entity and Image entity might contain the Anatomy entity.	For example, the Disease entity “二尖瓣后叶钙化 (posterior mitral valve leaflet calcification)” was recognized as the Anatomy entity “二尖瓣 (bicuspid),” and the Image entity “腹部彩超 (abdominal color doppler ultrasound)” was recognized as the Anatomy entity “腹部 (abdominal).”
On the self-annotated data set, entity nesting is more severe, the Disease entity, Examination entity, and Instrument entity might contain the Anatomy entity, and the Instrument entity might contain the Operation entity.	For example, the Disease entity “内踝骨折 (ankle fracture)” was recognized as the Anatomy entity “内踝 (medial malleolus),” the Examination entity “骨髓组织病理 (bone marrow histopathology)” was recognized as the Anatomy entity “骨髓 (bone marrow),” the Instrument entity “胸部支具 (chest brace)” was recognized as the Anatomy entity “胸(chest),” and the Instrument entity “左胸引流管 (left thoracic drainage tube)” was recognized as the Operation entity “左胸引流 (left thoracic drainage).”
(6) More entities with mixed representation	
Entity composition is more complex, mixed representations occur more often.	The Medicine entity “奥沙利铂 (乐沙定) (Oxaliplatin [Eloxatin])” on the Yidu-S4K data set was recognized as “奥沙利铂 (Oxaliplatin)” and “乐沙定 (Eloxatin),” respectively, the Disease entity “CD5[h]+弥漫大B细胞淋巴瘤 (白血病期)” on the self-annotated data set was recognized as “CD” and “弥漫大B细胞淋巴瘤 (白血病期) (diffuse large B-cell lymphoma [Leukemia stage]),” and the Examination entity “肥达、外斐反应 (Widal, well-felix reaction)” on the self-annotated data set was recognized as “肥达 (Widal)” and “外斐反应 (well-felix reaction),” respectively.
(7) Insufficient entity training data	
In the case of insufficient training samples, the machine may provide inadequate training for entities, so that the machine cannot fully learn the features of such entities, failing to recognize many entities.	On the self-annotated data set, the number of Instrument entities was less than that of other categories (Table 2), accounting for only 1.52% of the total, those entities might never appear in the training data set, such as “针筒 (syringe),” “微导管 (microtubule),” “550px碳钢钻头 (550px carbon steel drill bit),” etc.

[a] PLT: platelet count.

[b] NEUT: neutrophil count.

[c] CAE: carcinoembryonic antigen.

[d] PET-CT: positron emission tomography-computed tomography.

[e] AFP: alpha fetoprotein.

[f] HCG: human chorionic gonadotropin.

[g] VP: etoposide.

[h] CD5: a differentiation antigen, cluster of differentiation 5.

Discussion

Principal Findings

In this study, we developed a Chinese CNER method based on multisemantic features. The method extracted the semantic features of text using the RoBERTa-wwm model after dynamic fusion, extracted the fine-grained semantic features of Chinese characters by 1D CNN, and converted Chinese characters into square images to extract the image features of the simplified Chinese characters from another modality by 2D CNN. We conducted a series of experiments to evaluate the model's performance on the Yidu-S4K data set and self-annotated data set; the results showed that the F_1-scores of the proposed model in this study were 89.28% and 84.61% on the 2 data sets, respectively. The model showed a higher and more stable performance in all experiments and could help recognize entities in most categories. Furthermore, its migrative property and adaptability to different data were acceptable. We also demonstrated that multisemantic features were effective through 2 ablation experiences and analyzed the error cases of NER, which might provide a basis for subsequent studies and standardization of the corpus.

Compared with ensemble models, for the BiLSTM-CRF model, the representation information of characters was obtained with the help of a vector look-up table. However, the information obtained by this method was too simple to excavate the text's semantic meaning or solve problems such as the polysemy of words. Hence, the model did not perform well. Kong et al [69] constructed a multilayer CNN to obtain short-term and long-term contextual information, and the attention mechanism was used to calculate the weight distribution in each hidden layer so that the features of each coding layer could be fully extracted and used to improve the entity recognition performance. However, this model required numerous radical and dictionary features to complete the semantic supplement of the context. Li et al [70] proposed an ELMo-Lattice-LSTM-CRF model. The ELMo word dynamic representation model could learn complicated word features and the context-based changes of these features, while the lattice structure provided extra entity boundaries and other semantic information for CNER of EMRs through the Word2Vec model and dictionaries. Li et al [41] proposed an ELMo-BiLSTM-CRF model that improved the semantic recognition ability of the machine for text. It reduced problems, such as word polysemy, when compared with the BiLSTM-CRF model and reduced the computational complexity of the lattice structure compared with the ELMo-Lattice-LSTM-CRF model. Moreover, this model could fully use contextual information by replacing LSTM with BiLSTM. Wan et al [71] fine-tuned the ELMo model based on EMRs to achieve embedding for domain-specific text and then used a transformer as an encoder to alleviate the long context–dependent problems and finally achieved CNER through CRF decoding. Wang et al [72] proposed a model for NER based on the LSTM-CRF model by storing and merging characters, words, and other features. However, as the text embedding process of this method is more complicated, it is necessary to create dictionaries of characters and words to obtain multigranularity text features at first and then store and merge the obtained features using a tree structure to achieve text embedding. These methods have achieved a few good results, but our proposed method is still competitive and has the best performance among all the models, as shown in Table 4.

Compared with PLMs related to BERT, both the BERT-BiLSTM-CRF and BiLSTM-CRF models could obtain word-level vector representations. However, the word-level vector obtained by BERT contained rich contextual characteristics, including morphology, syntax, semantics, location, and other important semantic information, which can directly improve the task performance by replacing the lattice structure and complicated text representation methods in Table 4, such as dictionaries of characters and words. Compared with BERT, RoBERTa-wwm used more data for pretraining, and the dynamic wwm allows itself to flexibly learn word-level representation information, which compensates for the shortcomings that BERT can only obtain character-level representation. Thus, richer word-based text representation information could be obtained. Combined with the experimental results in Table 4, the RoBERTa-wwm-BiLSTM-CRF model, without introducing features, outperformed the other ensemble models. Therefore, using the PLM RoBERTa-wwm with a whole word mask can effectively improve the Chinese CNER performance, thus avoiding the use of complex text embedding and feature embedding methods.

In addition, 2 ablation experiments showed that different features and means lead to different degrees of improvement in the semantic comprehension ability of the model. Multisemantic features could help the machine to obtain richer semantic information, whereas dynamic fusion could fully recognize and used the representation information so that the model performance could be comprehensively improved. Considering the heterogeneity among data, using 1 method alone or both methods may affect the generalization ability of the model. In this study, the model combining the fine-grained semantic features and image features of Chinese characters and dynamic fusion might not show the best performance. However, it was more universal and could maintain the performance at a relatively high level compared with other experimental models. Furthermore, introducing more feature engineering was conducive to fully mining the semantic information of text connotation with the help of fine-grained semantic information contained in Chinese characters and improving the performance of the model on different data sets through the cross-complementarity of different features in a relatively stable manner.

To reduce the error rate of entity recognition, specifically for human-caused errors, we could try to avoid these problems by further improving the annotation quality. For the data special characteristics or data defects, the errors might be reduced by medical knowledge, medical dictionaries, and some rules, regardless of the lack of training data.

Limitations and Future Work

The limitation of this study was that the ratio of the 6 entity types on the Yidu-S4K data set did not exactly follow 7:3, such that the ratio of the training set to test set for disease entities is approximately 0.7610:0.2390; the ratio of the training set to test set for medicine entities is approximately 0.7898:0.2102; and the ratio of the training set to test set for all entities is approximately 0.7463:0.2537. The unbalanced data of different entity types in the training and test sets caused a performance bias. Although the ratio of the training set to the test set of the EMRs was 7:3, we could not ensure that the number of entities of each type in each EMR in the training set and test set remained at a similar ratio.

In the future, we may focus on the recognition of a specific entity type in EMRs to improve the CNER performance. In addition, we will incorporate other prior medical knowledge or assign different weights to the Chinese character semantic features and image features, such as using the attention mechanism to capture important features, to improve the performance of the model.

Conclusions

This study proposes a Chinese CNER method to learn a semantics-enriched representation of Chinese character features in EMRs to enhance the specificity and diversity of feature representations. The results showed that the model had state-of-the-art performance on 2 Chinese CNER data sets compared with existing models. We demonstrated that multisemantic features could provide richer and more fine-grained semantic information for Chinese CNER through the cross-complementarity of different semantic features. This enabled the model to learn a better feature representation and improve its generalization ability.

Acknowledgments

The authors would like to thank the YiduCloud for providing the Yidu-S4K corpora. This work was supported by the Science and Technology Innovation 2030 "New Generation Artificial Intelligence" major project "Research on the construction and application of knowledge system for Medical Artificial intelligence Services," China (grant 2020AAA0104901).

Conflicts of Interest

None declared.

References

1. Shen L, Li Q, Wang W, Zhu L, Zhao Q, Nie Y, et al. Treatment patterns and direct medical costs of metastatic colorectal cancer patients: a retrospective study of electronic medical records from urban China. J Med Econ 2020 May;23(5):456-463. [doi: 10.1080/13696998.2020.1717500] [Medline: 31950863]
2. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008:128-144. [Medline: 18660887]
3. Sang S, Yang Z, Liu X, Wang L, Lin H, Wang J, et al. GrEDeL: a knowledge graph embedding based method for drug discovery from biomedical literatures. IEEE Access 2018 Dec 12;7:8404-8415 [FREE Full text] [doi: 10.1109/access.2018.2886311]
4. Dogan RI, Lu Z. An improved corpus of disease mentions in PubMed citations. In: Proceedings of the 2012 Workshop on Biomedical Natural Language Processing. 2012 Presented at: BioNLP '12; June 8, 2012; Montreal, Canada p. 91-99 URL: https://dl.acm.org/doi/10.5555/2391123.2391135
5. Saha S, Ekbal A, Sikdar UK. Named entity recognition and classification in biomedical text using classifier ensemble. Int J Data Min Bioinform 2015;11(4):365-391. [doi: 10.1504/ijdmb.2015.067954] [Medline: 26336665]
6. Shao Y, Hardmeier C, Tiedemann J, Nivre J. Character-based joint segmentation and POS tagging for Chinese using bidirectional RNN-CRF. In: Proceedings of the 8th International Joint Conference on Natural Language Processing. 2017 Presented at: IJCNLP' 17; November 27–December 1, 2017; Taipei, Taiwan p. 173-183 URL: https://aclanthology.org/I17-1018.pdf
7. Peng H, Cambria E, Zou X. Radical-based hierarchical embeddings for Chinese sentiment analysis at sentence level. In: Proceedings of the 30th International Florida Artificial Intelligence Research Society Conference. 2017 Presented at: FLAIRS '17; May 22–24, 2017; Marco Island, FL, USA p. 347-352 URL: https://sentic.net/radical-embeddings-for-chinese-sentiment-analysis.pdf
8. Shi X, Zhai J, Yang X, Xie Z, Liu C. Radical embedding: delving deeper to Chinese radicals. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. 2015 Presented at: IJCNLP '15; July 26-31, 2015; Beijing, China p. 594-598 URL: https://aclanthology.org/P15-2098.pdf [doi: 10.3115/v1/p15-2098]
9. Zhang WY. Contrastive Studies on English and Chinese vocabulary. Shanghai, China: Shanghai Foreign Language Education Press; 2010.
10. Liu JN. A study on the structure of Chinese characters. Jilin University. 2011. URL: https://kns.cnki.net/kcms2/article/abstract?v=3uoqIhG8C447WN1SO36whHG-SvTYjkCc7dJWN_daf9c2-IbmsiYfKhSrIbg5xu9MmzfA5stc78MC2nDoIg7_tb-KveYJIGpg&uniplatform=NZKPT [accessed 2023-04-17]
11. Tsuruoka Y, Tsujii J. Improving the performance of dictionary-based approaches in protein name recognition. J Biomed Inform 2004 Dec;37(6):461-470 [FREE Full text] [doi: 10.1016/j.jbi.2004.08.003]

12. Song M, Yu H, Han WS. Developing a hybrid dictionary-based bio-entity recognition technique. BMC Med Inform Decis Mak 2015;15(Suppl 1):S9 [FREE Full text] [doi: 10.1186/1472-6947-15-S1-S9] [Medline: 26043907]
13. Kraus S, Blake C, West SL. Information extraction from medical notes. Medinfo. 2007. URL: https://ils.unc.edu/phr/files/KrausBlakeWestMEDINFO2007.pdf [accessed 2022-05-15]
14. Tsai RT, Sung CL, Dai HJ, Hung HC, Sung TY, Hsu WL. NERBio: using selected word conjunctions, term normalization, and global patterns to improve biomedical named entity recognition. BMC Bioinformatics 2006 Dec 18;7 Suppl 5(Suppl 5):S11 [FREE Full text] [doi: 10.1186/1471-2105-7-S5-S11] [Medline: 17254295]
15. Khalifa M, Shaalan K. Character convolutions for arabic named entity recognition with long short-term memory networks. Comput Speech Lang 2019 Nov;58(C):335-346 [FREE Full text] [doi: 10.1016/j.csl.2019.05.003]
16. Na SH, Kim H, Min J, Kim K. Improving LSTM CRFs using character-based compositions for Korean named entity recognition. Comput Speech Lang 2019 Mar;54:106-121 [FREE Full text] [doi: 10.1016/j.csl.2018.09.005]
17. Yu H, An J, Yoon J, Kim H, Ko Y. Simple methods to overcome the limitations of general word representations in natural language processing tasks. Comput Speech Lang 2020 Jan;59:91-113 [FREE Full text] [doi: 10.1016/j.csl.2019.04.009]
18. Saha SK, Sarkar S, Mitra P. Feature selection techniques for maximum entropy based biomedical named entity recognition. J Biomed Inform 2009 Oct;42(5):905-911 [FREE Full text] [doi: 10.1016/j.jbi.2008.12.012] [Medline: 19535010]
19. Lee KJ, Hwang YS, Kim S, Rim HC. Biomedical named entity recognition using two-phase model based on SVMs. J Biomed Inform 2004 Dec;37(6):436-447 [FREE Full text] [doi: 10.1016/j.jbi.2004.08.012] [Medline: 15542017]
20. Ju Z, Wang J, Zhu F. Named entity recognition from biomedical text using SVM. In: Proceedings of the 5th International Conference on Bioinformatics and Biomedical Engineering. 2011 Presented at: ICBBE '11; May 10-12, 2011; Wuhan, China p. 1-4 URL: https://ieeexplore.ieee.org/document/5779984 [doi: 10.1109/icbbe.2011.5779984]
21. Shen D, Zhang JJ, Zhou G, Su J, Tan CL. Effective adaptation of hidden Markov model-based named entity recognizer for biomedical domain. In: Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine. 2003 Presented at: BioMed '03; July 11, 2003; Sapporo, Japan p. 49-56. [doi: 10.3115/1118958.1118965]
22. Zhou G, Su J. Named entity recognition using an HMM-based chunk tagger. In: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. 2002 Presented at: ACL '02; July 7-12, 2002; Philadelphia, PA, USA p. 473-480 URL: https://dl.acm.org/doi/abs/10.3115/1073083.1073163 [doi: 10.3115/1073083.1073163]
23. Liu J, Huang M, Zhu X. Recognizing biomedical named entities using skip-chain conditional random fields. In: Proceedings of the 2010 Workshop on Biomedical Natural Language Processing. 2010 Presented at: BioNLP '10; July 15, 2010; Uppsala, Sweden p. 10-18 URL: https://dl.acm.org/doi/pdf/10.5555/1869961.1869963
24. Skeppstedt M, Kvist M, Nilsson GH, Dalianis H. Automatic recognition of disorders, findings, pharmaceuticals and body structures from clinical text: an annotation and machine learning study. J Biomed Inform 2014 Jun;49:148-158 [FREE Full text] [doi: 10.1016/j.jbi.2014.01.012] [Medline: 24508177]
25. Gorinski PJ, Wu H, Grover C, Tobin R, Talbot C, Whalley H, et al. Named entity recognition for electronic health records: a comparison of rule-based and machine learning approaches. arXiv. Preprint posted online on June 5, 2019 2023. [doi: 10.48550/arXiv.1903.03985]
26. Suárez-Paniagua V, Rivera Zavala RM, Segura-Bedmar I, Martínez P. A two-stage deep learning approach for extracting entities and relationships from medical texts. J Biomed Inform 2019 Nov;99:103285 [FREE Full text] [doi: 10.1016/j.jbi.2019.103285] [Medline: 31546016]
27. Cocos A, Fiks AG, Masino AJ. Deep learning for pharmacovigilance: recurrent neural network architectures for labeling adverse drug reactions in Twitter posts. J Am Med Inform Assoc 2017 Jul 01;24(4):813-821 [FREE Full text] [doi: 10.1093/jamia/ocw180] [Medline: 28339747]
28. Wang Q, Zhou Y, Ruan T, Gao D, Xia Y, He P. Incorporating dictionaries into deep neural networks for the Chinese clinical named entity recognition. J Biomed Inform 2019 Apr;92:103133 [FREE Full text] [doi: 10.1016/j.jbi.2019.103133] [Medline: 30818005]
29. Bengio Y, LeCun Y. Convolutional networks for images, speech, and time series. In: Arbib MA, editor. The Handbook of Brain Theory and Neural Networks. Cambridge, MA, USA: MIT Press; Oct 1998:255-258.
30. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015 Jan;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
31. Sundermeyer M, Schlüter R, Ney H. LSTM neural networks for language modeling. In: Proceedings of the 13th Annual Conference of the International Speech Communication Association. 2012 Presented at: INTERSPEECH '12; September 9-13, 2012; Portland, OR, USA p. 194-197 URL: https://www.isca-speech.org/archive/pdfs/interspeech_2012/sundermeyer12_interspeech.pdf [doi: 10.21437/interspeech.2012-65]
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. In: Proceedings of the Workshop at 1st International Conference on Learning Representations. 2013 Presented at: ICLR '13; May 2-4, 2013; Scottsdale, AZ, USA URL: https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf
33. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics:

Human Language Technologies. 2019 Presented at: NAACL '19; June 2-7, 2019; Minneapolis, Minnesota p. 4171-4186 URL: https://aclanthology.org/N19-1423.pdf

34. Ji B, Li S, Yu J, Liu R, Tang JT, Yu J, et al. A BiLSTM-CRF method to chinese electronic medical record named entity recognition. In: Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence. 2018 Presented at: ACAI '18; December 21-23, 2018; Sanya, China p. 1-6 URL: https://dl.acm.org/doi/proceedings/10.1145/3302425 [doi: 10.1145/3302425.3302465]
35. Huang Z, Xu W, Yu K. Bidirectional LSTM-CRF models for sequence tagging. arXiv. Preprint posted online on August 9, 2015 2023. [doi: 10.48550/arXiv.1508.01991]
36. Gridach M. Character-level neural network for biomedical named entity recognition. J Biomed Inform 2017 Jun;70:85-91 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.002] [Medline: 28502909]
37. Wei H, Gao M, Zhou A, Chen F, Qu W, Wang C, et al. Named entity recognition from biomedical texts using a fusion attention-based BiLSTM-CRF. IEEE Access 2019 Jun 04;7:73627-73636 [FREE Full text] [doi: 10.1109/access.2019.2920734]
38. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 6th International Conference on Neural Information Processing Systems. 2013 Presented at: NIPS'13; December 5-10, 2013; Lake Tahoe, NV, USA p. 3111-3119 URL: https://dl.acm.org/doi/10.5555/2999792.2999959
39. Si Y, Wang J, Xu H, Roberts K. Enhancing clinical concept extraction with contextual embeddings. J Am Med Inform Assoc 2019 Nov 01;26(11):1297-1304 [FREE Full text] [doi: 10.1093/jamia/ocz096] [Medline: 31265066]
40. Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, et al. Deep contextualized word representations. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018 Presented at: NAACL '18; June 1-6, 2018; New Orleans, LA, USA p. 2227-2237 URL: https://aclanthology.org/N18-1202.pdf [doi: 10.18653/v1/n18-1202]
41. Li N, Luo L, Ding Z, Song Y, Yang Z, Lin H. Improving Chinese clinical named entity recognition using stroke ELMo and transfer learning. In: Proceedings of the 2019 Evaluation Tasks at the China Conference on Knowledge Graph and Semantic Computing. 2019 Presented at: CCKS-Tasks '19; August 24-27, 2019; Hangzhou, China. [doi: 10.1007/978-981-15-1956-7_14]
42. Tutubalina E, Nikolenko S. Combination of deep recurrent neural networks and conditional random fields for extracting adverse drug reactions from user reviews. J Healthc Eng 2017;2017:9451342 [FREE Full text] [doi: 10.1155/2017/9451342] [Medline: 29177027]
43. Giorgi J, Bader GD. Towards reliable named entity recognition in the biomedical domain. Bioinformatics 2020 Jan 01;36(1):280-286 [FREE Full text] [doi: 10.1093/bioinformatics/btz504] [Medline: 31218364]
44. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
45. Kim YM, Lee TH. Korean clinical entity recognition from diagnosis text using BERT. BMC Med Inform Decis Mak 2020 Sep 30;20(Suppl 7):242 [FREE Full text] [doi: 10.1186/s12911-020-01241-8] [Medline: 32998724]
46. Wang Y, Ananiadou S, Tsujii J. Improving clinical named entity recognition in Chinese using the graphical and phonetic feature. BMC Med Inform Decis Mak 2019 Dec 23;19(Suppl 7):273 [FREE Full text] [doi: 10.1186/s12911-019-0980-z] [Medline: 31865903]
47. Yin M, Mou C, Xiong K, Ren J. Chinese clinical named entity recognition with radical-level feature and self-attention mechanism. J Biomed Inform 2019 Oct;98:103289 [FREE Full text] [doi: 10.1016/j.jbi.2019.103289] [Medline: 31541715]
48. Li Y, Du G, Xiang Y, Li S, Ma L, Shao D, et al. Towards Chinese clinical named entity recognition by dynamic embedding using domain-specific knowledge. J Biomed Inform 2020 Jun;106:103435 [FREE Full text] [doi: 10.1016/j.jbi.2020.103435] [Medline: 32360988]
49. Dong C, Zhang J, Zong C, Hattori M, Di H. Character-based LSTM-CRF with radical-level features for chinese named entity recognition. In: Proceedings of the 5th CCF Conference on Natural Language Processing and Chinese Computing, and 24th International Conference on Computer Processing of Oriental Languages. 2016 Presented at: NLPCC/ICCPOL '16; December 2-6, 2016; Kunming, China p. 2016 URL: https://link.springer.com/chapter/10.1007/978-3-319-50496-4_20 [doi: 10.1007/978-3-319-50496-4_20]
50. Qiu J, Wang Q, Zhou Y, Ruan T, Gao J. Fast and accurate recognition of Chinese clinical named entities with residual dilated convolutions. In: Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine. 2018 Presented at: BIBM '18; December 3-6, 2018; Madrid, Spain p. 935-942. [doi: 10.1109/bibm.2018.8621360]
51. Yang X, Zou L. A conditional random fields approach to clinical name entity recognition. In: Proceedings of the Evaluation Tasks at the China Conference on Knowledge Graph and Semantic Computing. 2018 Presented at: CCKS '18; August 14-17, 2018; Tianjin, China p. 1-6 URL: https://ceur-ws.org/Vol-2242/paper01.pdf
52. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online on July 26, 2019 2023. [doi: 10.48550/arXiv.1907.11692]

53. Cui Y, Che W, Liu T, Qin B, Yang Z. Pre-training with whole word masking for Chinese BERT. IEEE/ACM Trans Audio Speech Lang Process 2021;29:3504-3514 [FREE Full text] [doi: 10.1109/taslp.2021.3124365]
54. Crane K, Weischedel C, Wardetzky M. The heat method for distance computation. Commun ACM 2017 Oct 24;60(11):90-99 [FREE Full text] [doi: 10.1145/3131280]
55. Yidu-S4k: Yiducloud structured 4K dataset. OpenKG. URL: http://openkg.cn/dataset/yidu-s4k [accessed 2019-06-26]
56. Electronic medical records Center. Iiyi. URL: https://bingli.iiyi.com/ [accessed 2022-08-17]
57. Python software foundation. Python. URL: https://www.python.org/ [accessed 2022-10-28]
58. The Linux Foundation. Linux. URL: https://pytorch.org/ [accessed 2022-10-28]
59. ZDIC. URL: https://www.zdic.net/ [accessed 2023-04-17]
60. Jawahar G, Sagot B, Seddah D. What does BERT learn about the structure of language? In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: ACL '19; July 28- August 10, 2019; Florence, Italy p. 3651-3657 URL: https://aclanthology.org/P19-1356.pdf [doi: 10.18653/v1/p19-1356]
61. Zhang YQ, Wang Y. Identifying named entities of Chinese electronic medical records based on RoBERTa-wwm dynamic fusion model. Data Anal Knowl Discov 2022;6(2/3):242-250 [FREE Full text] [doi: 10.11925/infotech.2096-3467.2021.0951]
62. You XD, Ge HJ, Han JM, Li YX, Lü XQ. Recognition of complex entities in weapons and equipment field. Acta Sci Natur Univ Pekinensis 2022;58(3):391-404 [FREE Full text] [doi: 10.13209/j.0479-8023.2021.118]
63. Packard JL, Boltz WG. The origin and early development of the Chinese writing system. Language 1996 Dec;72(4):801 [FREE Full text] [doi: 10.2307/416104]
64. Cui SG, Chen JH, Xiaohong L. Named entity recognition of Chinese electronic medical records based on a hybrid neural network and medical MC-BERT. J Univ Sci Technol China 2022;51(04):565-571 [FREE Full text]
65. Meng Y, Wu W, Wang F, Li X, Nie P, Yin F, et al. Glyce: glyph-vectors for Chinese character representations. In: Proceedings of the 33rd Conference on Neural Information Processing Systems. 2019 Presented at: NeurIPS '19; December 8-14, 2019; Vancouver, Canada URL: https://papers.nips.cc/paper_files/paper/2019/hash/452bf208bf901322968557227b8f6efe-Abstract.html
66. Graves A, Ndez S, Schmidhuber J. Bidirectional LSTM networks for improved phoneme classification and recognition. In: Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications. 2005 Presented at: ICANN '05; September 11-15, 2005; Warsaw, Poland p. 799-804 URL: https://link.springer.com/chapter/10.1007/11550907_126 [doi: 10.1007/11550907_126]
67. Lafferty JD, McCallum AK, Pereira FC. Conditional random fields: probabilistic models for segmenting and labeling sequence data. In: Proceedings of the 18th International Conference on Machine Learning. 2001 Presented at: ICML '01; June 28- July 1, 2001; San Francisco, CA, USA p. 282-289 URL: https://dl.acm.org/doi/10.5555/645530.655813
68. Viterbi A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans Inform Theory 1967 Apr;13(2):260-269 [FREE Full text] [doi: 10.1109/tit.1967.1054010]
69. Kong J, Zhang L, Jiang M, Liu T. Incorporating multi-level CNN and attention mechanism for Chinese clinical named entity recognition. J Biomed Inform 2021 Apr;116:103737 [FREE Full text] [doi: 10.1016/j.jbi.2021.103737] [Medline: 33737207]
70. Li Y, Wang X, Hui L, Zou L, Li H, Xu L, et al. Chinese clinical named entity recognition in electronic medical records: development of a lattice long short-term memory model with contextualized character representations. JMIR Med Inform 2020 Sep 04;8(9):e19848 [FREE Full text] [doi: 10.2196/19848] [Medline: 32885786]
71. Wan Q, Liu J, Wei L, Ji B. A self-attention based neural architecture for Chinese medical named entity recognition. Math Biosci Eng 2020 May 09;17(4):3498-3511 [FREE Full text] [doi: 10.3934/mbe.2020197] [Medline: 32987540]
72. Wang C, Wang H, Zhuang H, Li W, Han S, Zhang H, et al. Chinese medical named entity recognition based on multi-granularity semantic dictionary and multimodal tree. J Biomed Inform 2020 Nov;111:103583 [FREE Full text] [doi: 10.1016/j.jbi.2020.103583] [Medline: 33010427]

Abbreviations

BERT: Bidirectional Encoder Representation from Transformers
BiLSTM: Bidirectional Long Short-Term Memory
CNER: clinical named entity recognition
CNN: convolutional neural network
CRF: conditional random fields
ELMo: Embeddings from Language Models
EMR: electronic medical record
FN: false negative
FP: false positive
LSTM: long short-term memory
NER: named entity recognition
PLM: pretrained language model

RoBERTa: Robustly Optimized Bidirectional Encoder Representation from Transformers Pretraining Approach
TP: true positive
Word2Vec: Word to Vector
wwm: Whole Word Masking

Edited by G Eysenbach; submitted 01.12.22; peer-reviewed by C Yu, L Heryawan; comments to author 31.01.23; revised version received 18.02.23; accepted 31.03.23; published 10.05.23

Please cite as:
Wang W, Li X, Ren H, Gao D, Fang A
Chinese Clinical Named Entity Recognition From Electronic Medical Records Based on Multisemantic Features by Using Robustly Optimized Bidirectional Encoder Representation From Transformers Pretraining Approach Whole Word Masking and Convolutional Neural Networks: Model Development and Validation
JMIR Med Inform 2023;11:e44597
URL: https://medinform.jmir.org/2023/1/e44597
doi: 10.2196/44597
PMID: 37163343