Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 21(3): 4085–4103.
DOI: 10.3934/mbe.2024180
Received: 09 December 2023
Revised: 24 January 2024
Accepted: 01 February 2024
Published: 23 February 2024

Research article

Data governance and Gensini score automatic calculation for coronary angiography with deep-learning-based natural language extraction

Feng Li[1], Mingfeng Jiang[1], Hongzeng Xu[2], Yi Chen[1], Feng Chen[1], Wei Nie[1] and Li Wang[3,*]

[1] School of Information and Electronic Engineering, Zhejiang Gongshang University, Hangzhou 310018, China

[2] Department of Cardiology, The People's Hospital of Liaoning Province, Liaoning, Shenyang 110011, China

[3] College of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China

* **Correspondence:** Email: liwang2002@dlmu.edu.cn.

Abstract: With the widespread adoption of electronic health records, the amount of stored medical data has been increasing. Clinical data, often in the form of semi-structured or unstructured electronic medical records (EMRs), contains rich patient information. However, due to the use of natural language by physicians when composing these records, the effectiveness of traditional methods such as dictionaries, rule matching, and machine learning in the extraction of information from these unstructured texts falls short of clinical standards. In this paper, a novel deep-learning-based natural language extraction method is proposed to overcome current shortcomings in data governance and Gensini score automatic calculation in coronary angiography. A pre-trained model called bidirectional encoder representation from transformers (BERT) with strong text feature representation capabilities is employed as the feature representation layer. It is combined with bidirectional long short-term memory (BiLSTM) and conditional random field (CRF) models to extract both global and local features from the text. The study included an evaluation of the model on a dataset from a hospital in China and it was compared with another model to validate its practical advantages. Hence, the BiLSTM-CRF model was employed to automatically extract relevant coronary angiogram information from EMR texts. The achieved F1 score was 91.19, which is approximately 0.87 higher than the BERT-BiLSTM-CRF model.

Keywords: electronic health records; coronary angiography; Gensini score; deep learning

1. Introduction

Named entity recognition (NER) refers to the identification of specific entities in the text, such as the names of individuals, locations and organizations. This constitutes a significant aspect of information

extraction [1]. The current studies show that medical information provided by electronic medical records (EMRs) is more complete and faster to retrieve than traditional paper records [2]. Nowdays, EMRs are becoming the main source of medical information about patients [3]. Concurrently, the value of medical information has become increasingly significant. Due to the unstructured nature of these records, extracting information from clinical notes is quite challenging. Entity extraction from electronic health records forms the foundation for studying patient conditions. For healthcare professionals, the deep exploration of clinical data for research purposes is not straightforward, and there is a need for tools to facilitate this process. Deep learning can be employed for the automated analysis of coronary angiographic images, reducing manual interventions and subjective judgments. By training deep learning models, the automated identification and quantitative assessment of coronary artery lesions, degrees of narrowing, and hemodynamics can be determined, offering rapid and accurate diagnostic results.

Surgical information constitutes an essential component of clinical data, often stored in unstructured form within electronic health records (EHRs) [4]. Deep learning models can serve as assistive tools for physicians, providing interpretation suggestions and diagnostic opinions for coronary angiographic images [5]. These models can automatically identify and label abnormal areas, assisting physicians in the determination of lesion locations, severity, as well as suggesting potential treatment strategies. By training on a large dataset of coronary angiographic images and clinical data, predictive models can be established to assess a patient's risk of coronary heart disease. Such models aid in the early identification of high-risk patients, enabling preventative measures and reducing the occurrence of cardiovascular events. Deep learning models can uncover previously neglected or hard-to-discern features and patterns when analyzing coronary angiographic images. These new insights deepen the understanding of coronary artery disease, drive medical research progress, and offer new leads for disease diagnosis and treatment [6].

In summary, the roles of deep learning in the extraction of coronary angiographic images include automated analysis, diagnostic assistance, risk assessment, and knowledge discovery. This can enhance early diagnosis rates, treatment accuracy, and patient management efficiency for coronary heart disease. However, the clinical application of deep learning models requires thorough validation and further research to ensure accuracy, reliability, and safety.

The purpose of this study was to design a system for the extraction of coronary angiography results based on EMR text. Several machine learning models were compared, and the model combining bidirectional long short-term memory (BiLSTM) and conditional random field (CRF) algorithms yielded the best performance on the task of identifying risk factors. Considering the performance of existing risk factor extraction systems, our solution provides an efficient way to extract coronary angiography risk factors from EMRs. Our model achieved an accuracy of 91.19%, and it was obtained within 200 ms when applied. We have applied deep learning methods for coronary angiography surgery-related extraction for the first time and compared the results with those of traditional methods; we have achieved both accuracy and speed improvements. And the datasets we used were all real data from hospitals, allowing the models we ran to be applied in the medical field.

The Gensini score was used to assess the severity of coronary artery disease in patients. By evaluating scores and analyzing data from a large number of patients, the effectiveness and safety of new treatments, drugs, or interventions, as well as factors related to patient selection and outcomes, can be assessed. After implementing deep learning algorithms, automated calculation of the Gensini score can

be realized. As compared to previous manual calculations, where doctors had to process each patient's data and perform calculations, our automated process significantly reduces the workload for medical professionals and provides more accurate results.

2. Related work

NER is commonly regarded as a sequence labeling task in machine learning, and it can be realized through the application of various methods, including rule-based and dictionary-based methods, statistical machine learning methods, and deep learning methods. Rule-based and dictionary-based approaches involve the construction of templates or dictionaries by experts and utilization of matching techniques to process the text. These methods are limited in terms of their labor-intensiveness, scalability to other domains, and adaptation to changing data. Machine learning methods for NER include hidden Markov models [7], support vector machines [8], CRFs [9], and more. Deep learning methods leverage large datasets to automatically learn features. These methods optimize model parameters during training to enhance learning efficiency and accuracy, avoiding the subjectivity and randomness associated with manual feature selection. Long short-term memory (LSTM), proposed by Hochreiter and others [10], is a widely used type of deep learning model that addresses the long-term dependency issue of traditional recurrent neural networks (RNNs) through the use of "gates" and cell concepts. In 2018, significant advancements were made in natural language processing (NLP) with the introduction of models like bidirectional encoder representations from transformers (BERT) [11], ELMo [12], and OpenAI GPT. BERT, a pre-trained language model, achieved state-of-the-art results on various NLP subtasks and improved the performance of 11 downstream NLP tasks [13].

CRF and BiLSTM networks with CRFs (BiLSTM-CRFs) are widely used methods for Chinese NER tasks. While these methods exhibit reliability, they mostly focus on English electronic health record (EHR) environments. Recognizing named entities in Chinese text remains challenging due to the lack of clear word boundaries and the complexity of Chinese language structure.

For Chinese NER, Ouyang et al. proposed the concept of NER based on the BiLSTM neural network with additional features [14]. However, they did not incorporate CRF or consider the character information in words. Xiang proposed a Chinese NER method based on character-word mixed embedding [15]. They averaged the word embedding and character embedding when the word contains only one character. Yang and Gao proposed deep neural networks for medical entity recognition in Chinese online health consultations [16]. They utilized a BI-LSTM-CRF network as the basic architecture and concatenated character embedding and context word embedding to learn effective features.

Zhang et al. compared the efficiency of entity recognition models by using different feature templates and context windows in CRF learning and training, seeking the optimal electronic health record feature templates and context windows [17]. They found that the BiLSTM-CRF model outperformed traditional CRF-based methods on multiple datasets. Various researchers have experimented with the BiLSTM-CRF model on Chinese medical texts and electronic health records. Zhang et al. utilized CRFs and the BiLSTM-CRF to extract disease, body part, and treatment information from electronic health record datasets, finding that the latter performed better. Qiu et al. proposed the residual convolutional neural network CRF model to improve the training speed while achieving higher F1 scores than the BiLSTM-CRF model [18]. In conclusion, NER methods range from rule-based approaches to deep learning models based on the BiLSTM-CRF, which have demonstrated significant advancements

on various tasks, especially when applied to medical texts and electronic health records [19–23].

3. Materials and methods

In this section, we will provide a detailed description for the sources and composition of our dataset, as well as the deep learning models we choose to use.

3.1. Dataset

The experimental data used in this study was mainly taken from clinical records of nearly 20,000 patients from a famous hospital in Liaoning Province, China, primarily from the Department of Cardiology. After organizing the data, we selected one thousand cases for analysis. For each patient, the emergency room recorded a series of clinical notes. These records include the patient's outpatient number, surgery date, surgery start time, surgery end time, and coronary angiogram results. We extracted information specifically related to patients' coronary angiography procedures.

To extract information from the EMRs, we employed the BIO tagging method and deployed a medical annotation website by using the text annotation tool "doccano" [24]. Doccano supports three types of NLP tasks for text annotation: text classification, sequence labeling, and sequence-to-sequence tasks (such as text translation).

We categorized the results of patients' coronary angiography procedures into five labels: left anterior descending (LAD) artery, left circumflex (LCX) artery, right coronary artery (RCA), left main coronary artery (LMCA) and SupportBrand, which correspond to a support stent. We eventually obtained approximately nine hundred and seventy processed cases.

Among the 970 subsets of data used, we processed 717 subsets for the LAD, 650 subsets for the LCX, 102 subsets for the LMCA, 681 subsets for the RCA, and subsets for the SupportBrand. And we set the ratio of training, testing, and validation sets for the dataset to 8:1:1.

Table 1 illustrates examples of extracted coronary angiography procedural results from the original EMR text shown in Figure 1. The entities in Table 1 correspond to annotations in the text of Figure 1.

In Table 1, the entities are associated with their corresponding entity types, starting positions, and ending positions.

The left coronary artery originates from the left coronary sinus of the aortic root and travels approximately 0.5–2 cm; it is referred to as the LMCA. It then branches into the LAD and LCX arteries. The LAD courses through the anterior interventricular groove, located at the front of the heart. The circumflex artery runs along the lateral aspect of the heart. In some individuals, an intermediate branch may arise between the LAD and the LCX artery. The LAD gives rise to branches including septal branches and diagonal branches. The LCX artery's branches include obtuse marginal branches and left atrial branches. In individuals with a left dominant circulation, the LCX artery may also give rise to a left posterior descending branch.

The RCA originates from the right coronary sinus of the aortic root and courses within the atrioventricular groove, situated on the right side of the heart. It wraps around the inferior surface of the heart and reaches the diaphragmatic surface. The RCA gives rise to the conus branches, right atrial branches, right ventricular branches, acute marginal branches, left posterior ventricular branches, and posterior descending branches. Through these vessels and their branches, the heart receives blood supply to various regions.

Table 1. Examples of coronary angiography procedural results as extracted from EMR text.

Character entity	Starting Position	Ending Position	Entity Type
Left main trunk: 40% stenosis in the middle to far segment	113	126	LMCA
Anterior descending branch: 50% stenosis in the proximal to middle segment, blood flow TIMI3 level	127	150	LAD
Circumflex branch: Middle segment 80% stenosis (small), blood flow TIMI3 level	150	174	LCX
Right coronary artery: Middle segment 100% occlusion, visible thrombus shadow, blood flow TIMI0 level	175	201	RCA
Implantation of 3.0mm × 10mm/10–12 atm Aili stent and Lepu stent	308	336	SupportBrand

The extracted data holds significant clinical research value and greatly assist in determination of the presence of cardiovascular diseases in patients.

3.2. Methods

NER [25] can be framed as a sequence labeling task that targets entities within clinical texts. The conversion of JSON-format annotations into the BIO format is a prerequisite for subsequent operations. In Chinese language corpora, character-based NER methods tend to be more effective than word-based approaches. Sentence labels were marked by following the BIO standard. Each character in the sentence was assigned a tag from the set including O, B-PER, I-PER, B-LOC, I-LOC, B-ORG and I-ORG. Here, "O" indicates no entity association, "B" signifies the start of an element in a fragment, and "I" denotes the middle position of an element within a fragment. Entities are represented as "B-X" and "I-X", while non-entities are represented as "O". For this task, there were five entity types, resulting in eleven labels: B-LAD, I-LAD, B-LCA, I-LCA, B-RCA, I-RCA, B-LMCA, I-LMCA, B-SupportBrand, I-SupportBrand, and O.

This study explored three medical NER methods: traditional string extraction, the BiLSTM-CRF model, and the pre-trained BERT character embedding combined with BiLSTM-CRF (BERT-BiLSTM-CRF) model. After annotating Chinese EMRs in JSON format and converting it to BIO format, features such as part of speech, radical, document type (DOC type), and character position (CHAR index) were selected as the CRF features. Character embeddings serve as input features for both the BiLSTM-CRF model and the BERT-BiLSTM-CRF model, and a dictionary was constructed.

患者今日行冠状动脉造影术+支架植入术.患者取仰卧位，常规消毒双侧腹股沟区皮肤，铺无菌巾，1%利多卡因局部浸润麻醉，Seldinger 技术穿刺右股动脉，成功后置入规格 6F 动脉鞘管。经鞘管置入 JR4.0、JL4.0 造影导管造影见:左主干：中-远段 40%狭窄，前降支：近-中段 50%狭窄，血流 TIMI3 级，回旋支：中段 80%狭窄（细小），血流 TIMI3 级，右冠：中段 100%闭塞，可见血栓影，血流 TIMI0 级。左室造影未做。支架植入术：推送 6FJR4.0 指引导管至右冠口，送 SIONblue 导丝至右冠远端。右冠中段：2.5mm*15mm/10atmApex 球囊预扩张→由远及近依次植入 3.0mm*10mm/10-12atm 爱立支架，3.0mm*24mm/10-12atm 乐普支架，残余狭窄 0%,血流 TIMI3 级退出指引导管，拔除桡动脉鞘，压迫器压迫术区，送返心内监护病房。术后血压 109/70mmHg。心率 126 次/分。术中应用肝素 7000u、普优克 10ml 造影剂碘克沙醇 130ml。↵
↵

The patient underwent coronary angiography and stent implantation today. The patient was placed in a supine position, and routine disinfection was performed on both inguinal regions. Sterile drapes were laid, and 1% lidocaine local anesthesia was administered. Using the Seldinger technique, the right femoral artery was punctured, and a 6F arterial sheath was successfully inserted. Through the sheath, JR4.0 and JL4.0 catheters were introduced for angiography: Left main artery: 40% stenosis in the mid to distal segment, left anterior descending artery: 50% stenosis in the proximal to mid segment, blood flow TIMI3 grade; Circumflex artery: 80% stenosis in the mid segment (small vessel), blood flow TIMI3 grade; Right coronary artery: 100% occlusion in the mid segment, visible thrombus, blood flow TIMI0 grade. Left ventriculography was not performed. Stent implantation: A 6F JR4.0 guiding catheter was advanced to the ostium of the right coronary artery, and a SIONblue guidewire was guided to the distal right coronary artery. In the mid segment of the right coronary artery: pre-dilation with a 2.5mm*15mm/10atm Apex balloon, followed by sequential implantation of a 3.0mm*10mm/10-12atm Xience stent and a 3.0mm*24mm/10-12atm Resolute stent from distal to proximal. Residual stenosis was 0%, blood flow TIMI3 grade. The guiding catheter was removed, the radial artery sheath was extracted, compression was applied to the access site, and the patient was transferred back to the cardiac care unit. Post-procedure blood pressure was 109/70mmHg, and heart rate was 126 beats per minute. During the operation, 7000u of heparin and 10ml of Puyuke contrast agent, 130ml of iodixanol, were used.↵

Figure 1. An example of the original text of coronary angiography procedural results.

3.2.1. Dictionary-based method

For dictionary-based methods, the adoption of a forward maximal matching algorithm is common. This algorithm consists of two main steps: it typically starts from the beginning of the string; then, the network selects the longest segment with a word length. If the sequence does not exceed the maximum word length, choose all sequences. First, check if this segment is in the dictionary. If it is, consider a separated word. If not, start reducing one character from the right. Then, check if there are shorter segments in the dictionary and repeat until only one word remains. After the segmentation, the sequence becomes the remaining part of the sequence. The "forward" in the method of forward maximum matching implies that the matching proceeds from front to back. The maximal means that the longer the word length we match, the better the result we can obtain, i.e., fewer words separated in this sentence means a better outcome.. We can treat terms like LMCA or LAD as our matching fields and automatically extract the desired content. However, we found that after attempting this approach, the accuracy is not very high. If there are texts that do not align with our string matches, it can lead to incorrect results. As a result, this method may introduce numerous errors. Nevertheless, in certain ordinary medical contexts, the extraction method based on string matching can still be applicable.

3.2.2. The CRF model

A CRF is a statistical modeling method commonly employed for sequence labeling tasks such as NER, part-of-speech tagging, and semantic role labeling. It is a discriminative model that builds a joint probability model of given input sequences and the corresponding label sequences to perform sequence labeling. The fundamental idea behind the CRF model is to establish a Markov random field model for label sequences (output sequences) given an observed sequence (input sequence). The CRF model defines feature functions to represent the relationships between input and label sequences, utilizing weight parameters associated with these feature functions for learning.

Specifically, the CRF model transforms sequence labeling problems into an optimization problem to allow the model to find a set of parameters. During the training phase, the model's parameters are estimated by maximizing the log-likelihood function of the training data. This is typically achieved by using iterative methods such as stochastic gradient descent or quasi-Newton methods. During the predictive phase, the learned parameters are used to choose the optimal label sequence as the output through the use of decoding algorithms.

In summary, the CRF model is a commonly used approach for sequence labeling tasks, effectively capturing correlations between input and output sequences. It has demonstrated strong performance on various NLP tasks.

Suppose that we have an observation sequence (input sequence) $X = \{x_1, x_2, \cdots, x_n\}$, along with the corresponding label sequence (output sequence) $Y = \{y_1, y_2, \cdots, y_n\}$. Given the observation sequence X, the probability of a CRF can be represented as follows:

$$P(Y|X) = \frac{1}{Z(x)} exp(\sum_{i=1}^{n} \sum_{k=1}^{K} \lambda_k * \phi_k(y_{i-1}, y_i, x, i)) \tag{3.1}$$

where X is the normalization factor, defined as the sum of probabilities of all possible label sequences:

$$Z(X) = \sum yexp(\sum_{i=1}^{n}\sum_{k=1}^{K}\lambda_k * \phi_k(y_{i-1}, y_i, x, i)) \tag{3.2}$$

where λ_k represents the weight parameter associated with the feature function ϕ_k, which needs to be learned during the training process.

During the predictive phase, we aim to find the optimal sequence of labels $\widehat{Y}$, which maximizes the probability of the CRF. An efficient optimal label sequence decoding technique entails the use of dynamic programming algorithms, such as the Viterbi algorithm.

3.2.3. The BiLSTM-CRF model

In Table 2, batch_size represents the number of data samples passed to the program for training in a single iteration, hidden_dim denotes the number of nodes in the hidden layer, embedding_dim indicates the number of nodes in the embedding layer, *lr* stands for the learning rate, and the Adam optimizer was chosen. The term dropout results signified a dropout rate of 0.5.

The BiLSTM-CRF model is a commonly used deep learning model for sequence labeling tasks. It combines the components of BiLSTM and the CRF to enhance the performance of sequence labeling tasks.

BiLSTM is a variant of RNNs that is suitable for processing sequential data. Unlike traditional RNNs, BiLSTM considers contextual information from both directions by incorporating two sets of LSTM units: one processes the sequence in the forward direction and the other in the reverse direction. This allows the model to capture information from both the preceding and subsequent elements at each time step.

During the predictive phase, the softmax layer is typically used to address multi-class classification problems, but it does not consider the dependencies in the relationships between labels during sequence labeling tasks. This can lead to syntactically incorrect label sequences. To address this, the CRF layer is utilized. In the CRF model's output, label types like I-X do not appear after O-labels, as I-X labels must follow B-X labels. By inputting the features learned by the BiLSTM network into the CRF layer, syntactic issues are avoided; also, recognition accuracy is enhanced due to feature diversity.

The BiLSTM-CRF model combines BiLSTM and CRF components. It employs BiLSTM to extract sequence features and then utilizes the CRF layer for labeling. Specifically, BiLSTM maps the input sequence at each time step into a high-dimensional feature space, and the CRF layer assigns labels to these features. The CRF layer utilizes the Viterbi algorithm to decode and identify the most probable label sequence.

In summary, the BiLSTM-CRF model is a deep learning architecture that integrates BiLSTM and CRF characteristics. It is used for sequence labeling tasks and effectively handles relationships between sequential data and labels.

Each word in the sentence is represented by a word vector containing both word embeddings and character embeddings. Word embeddings are usually pre-trained, while character embeddings are randomly initialized. All embeddings are adjusted throughout the training iterations. The input for the BiLSTM-CRF model consists of word embedding vectors, and the output is the predicted label for each word. The scores outputted by the BiLSTM layer for all labels are used as input for the CRF layer, where the label with the highest score in the sequence becomes our final predicted result.

Table 2. Information about the parameters of the BiLSTM-CRF model.

Parameter names	Parameter information
batch_size	32
hidden_dim	300
embedding_dim	300
lr	0.001
clip	5.0
optimizer	Adam
dropout	0.5

Figure 2 shows a BiLSTM+CRF layer model, wherein the BiLSTM, after adding the CRF layer to the output, can capture contextual information, enhancing the correlation between textual information. For sequence annotation problems, each word in each sentence has an annotation result. High-dimensional features are extracted for the *i*th word in a sentence. By learning the mapping from features to annotation results, probabilities for any label can be obtained. Based on these probabilities, the optimal sequence structure can be determined. The algorithm proceeds with the forward propagation of the BiLSTM+CRF model, followed by the forward and backward propagation of the CRF layer, and finally, the backward propagation of the BiLSTM+CRF model updates parameters to conclude the loop.

3.2.4. The BERT-BiLSTM-CRF model

The BERT-BiLSTM-CRF model is similar to the BiLSTM-CRF model, with the key difference lying in the input word vectors. The distinction is that the BERT model is incorporated as a feature representation layer within the BiLSTM model. This integration is utilized to generate word vectors that are input into the LSTM structure.

As shown in Figure 3, the model is basically consistent with the content shown in Figure 2, except that all of the BERT layers have been added in front of it. The features learned by each layer of the BERT model are not entirely the same. The lower layers of the BERT model mainly acquire phrase-level feature information, the middle layers primarily learn syntactic structural feature information, and the top layers capture the semantic information of the entire sentence. After undergoing BERT processing, contextualized word vectors are obtained, which is key in the handling of long-range dependency information in sentences.

3.2.5. Automated Gensini rating calculation

The Gensini score plays a role in the evaluation of the severity of coronary artery lesions in patients.

The Gensini score for coronary arteries divides the coronary artery system into 15 segments based on the segmental criteria recommended by the American Heart Association, as per the World Health Organization standards. It includes 1 segment for the LMCA with a coefficient of 5. The RCA was divided into 4 segments (proximal segment coefficient of 2.5, middle segment coefficient of 1.5, distal segment coefficient of 1.0, and right posterior descending branch coefficient of 1.0). The LAD artery was divided into 5 segments (proximal segment coefficient of 2.5, middle segment coefficient of 1.5,

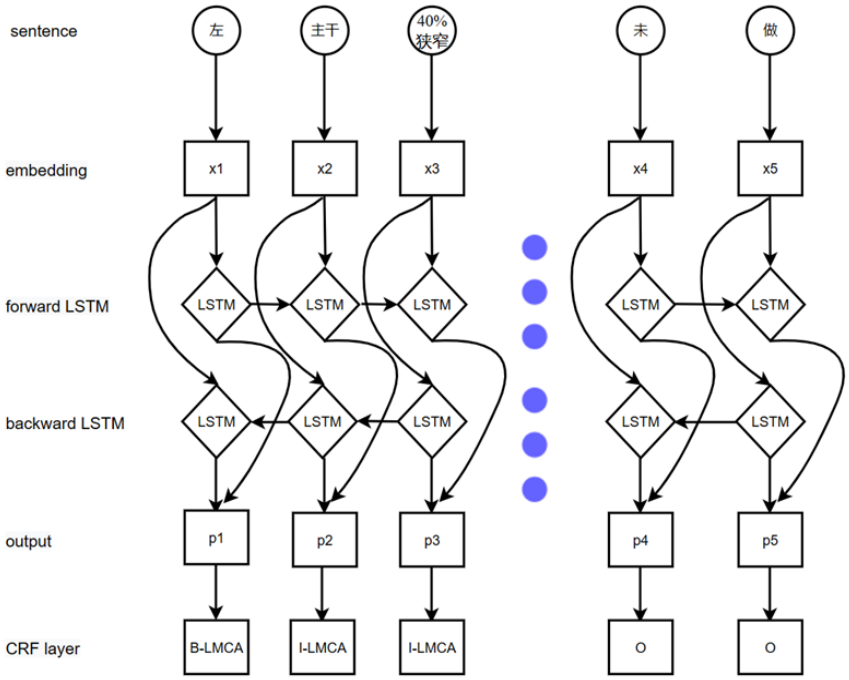

Figure 2. BiLSTM-CRF model diagram.

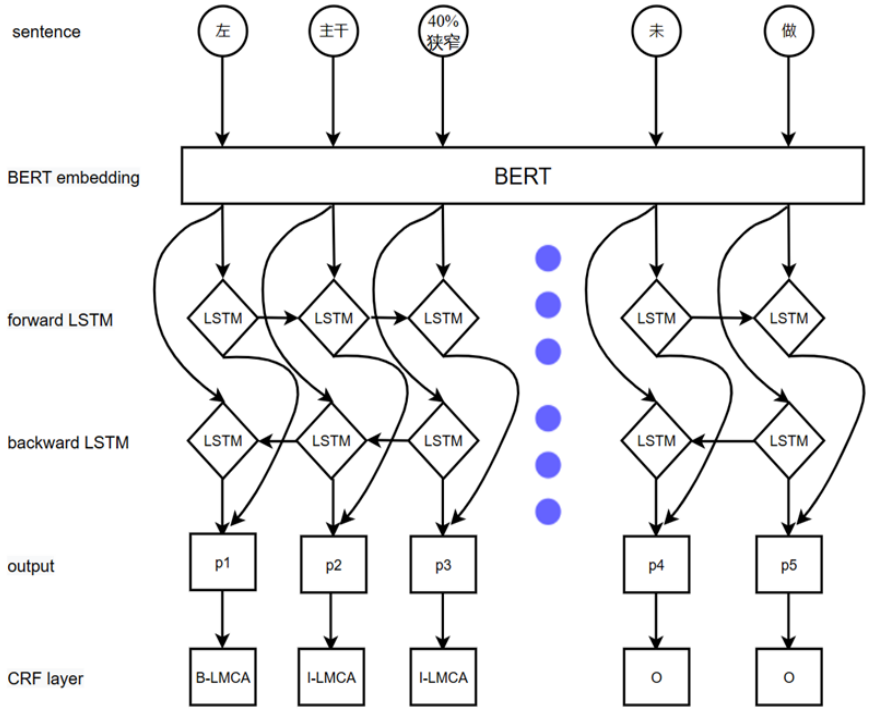

Figure 3. BERT-BiLSTM-CRF model diagram.

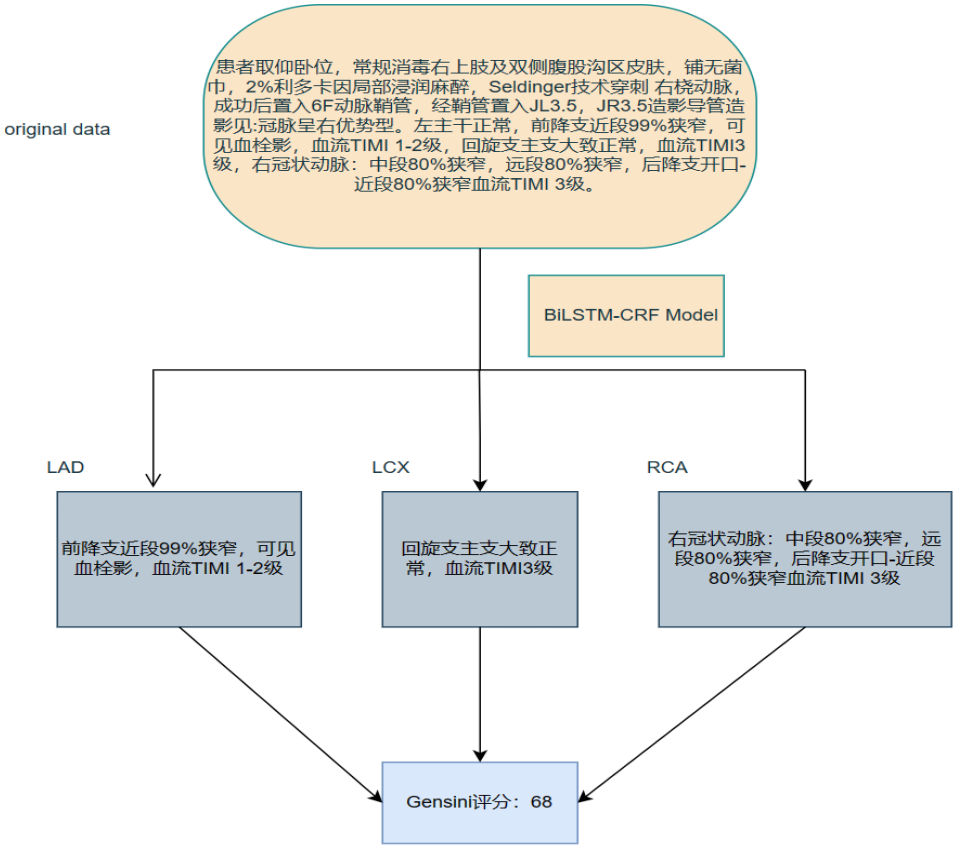

Figure 4. Gensini rating.

distal segment coefficient of 1.0, first diagonal branch coefficient of 1.0, and second diagonal branch coefficient of 0.5). The LCX artery was divided into 5 segments (proximal segment coefficient of 2.5, middle segment coefficient of 1.5, distal segment coefficient of 1.0, first obtuse marginal branch coefficient of 0.5, and second obtuse marginal branch coefficient of 0.5). The degree of stenosis is defined as follows: 0–25% stenosis scores 1 point; 26%–50% stenosis scores 2 points; 51%–75% stenosis scores 4 points; 76%–90% stenosis scores 8 points; 91%–99% stenosis scores 16 points; 100% stenosis scores 32 points. The Gensini score for each patient is calculated by summing the product of stenosis degree points and the respective coefficients. The highest score is realized when the LMCA is completely occluded (32 * 5 = 160), and a score of 0 is yielded when there is no coronary artery disease.

We can ascertain from Figure 4 that, after inputting the original data into our model, we can automatically extract the degree of narrowing and blood flow information from different parts of the text. Subsequently, by processing the obtained results through steps like string manipulation, we can automatically calculate the Gensini score that we need.

In Figure 4, the original data shows that the patient was placed in a supine position, with routine disinfection of the skin on the right upper limb and bilateral inguinal areas, sterile wipes applied, and local infiltration anesthesia with 2% lidocaine. The Seldinger technique was used to puncture the right

Table 3. The results of the BERT-BiLSTM-CRF model.

Tag Name/Result	Precision	Recall	F1 score
LAD	89.71%	93.85%	91.73
LCX	85.61%	90.15%	87.82
LMCA	70.37%	82.61%	76.00
RCA	86.33%	93.02%	89.55
SupportBrand	88.83%	98.87%	93.58
Total	86.99%	93.91%	90.32

radial artery, and a 6F arterial sheath was successfully inserted. JL3.5 was inserted through the sheath, and a right dominant coronary artery was observed in the imaging of the JR3.5 catheter. The left main artery was normal, with 99% stenosis in the proximal segment of the anterior descending branch and visible thrombus shadows. The blood flow was TIMI level 1–2, while the main branch of the circumflex branch was generally normal at a TIMI grade 3. The RCA had 80% stenosis in the middle segment, 80% stenosis in the distal segment, and 80% stenosis in the opening to the proximal segment of the posterior descending branch and a TIMI grade 3. The LAD artery exhibited 99% stenosis in the proximal segment of the anterior descending branch, with visible thrombus shadows and TIMI grade 1–2 blood flow. The LCX artery is the main branch of the circumflex branch, which is generally normal with blood flow at a TIMI 3. RCA exhibited 80% stenosis in the middle segment, 80% stenosis in the far segment, and TIMI grade 3 blood flow from the opening of the posterior descending branch to 80% stenosis in the near segment. The final Gensini score obtained was 68.

4. Experimental results

In this section, we will analyze the various results we have obtained through the model, as well as the time complexity and other information regarding the model.

4.1. Result of our models

The F1 score metric ranges from 0 to 1, with a higher value indicating better classification performance of the model. The F1 score model metric combines precision and recall, allowing for a more comprehensive assessment of the model's classification performance. When the F1 score model metric is close to 1, it indicates excellent classification performance, particularly that the model can accurately predict positive instances while capturing the majority of them. Conversely, when the FB1 is close to 0, it suggests poorer classification performance, potentially involving more false positives and false negatives. Model metrics have widespread applications in practical scenarios. For instance, in case of text classification tasks, the F1 score can be used to evaluate the model's performance on the task of classifying different categories, aiding in the selection of the most suitable model. In recommendation systems, the F1 score model metric can assess the accuracy and comprehensiveness of the model's recommendations for different users, optimizing the recommendation algorithm. In medical diagnosis, the FB1 can be employed to evaluate the accuracy and comprehensiveness of the model in diagnosis of various medical conditions, assisting doctors in making precise diagnoses.

Table 4. The results of the BiLSTM-CRF model

Tag Name/Result	Precision	Recall	F1 score
LAD	90.37%	93.85%	92.08
LCX	90.30%	91.67%	90.98
LMCA	70.83%	73.91%	72.34
RCA	90.37%	94.57%	92.42
SupportBrand	87.76%	97.18%	92.23
Total	88.78%	93.74%	91.19

$$Precision = \frac{TP}{TP + FP} \tag{4.1}$$

$$Recall = \frac{TP}{TP + FN} \tag{4.2}$$

$$F1 = \frac{2Precision * Recall}{Precision + Recall} \tag{4.3}$$

Here, TP, FP, and FN represent true positives, false positives, and false negatives, respectively. For multi-class classification, there are macro- and micro-average precision (P), recall (R), and F1-score (F) evaluation methods. In the macro-average method, each entity type is evaluated as positive, while other entity types are considered negative; then, the average score across all types is calculated. On the other hand, the micro-average method treats the evaluation of all entity types equally. It has been widely applied in the assessment of risk factor extraction systems in the i2b2 2014 challenge. The micro-average evaluation is believed to consider all possible error types and overlooks the imbalance in type distribution. In this context, we utilized the micro-average evaluation metrics, and the P, R, and F1 score values in the experiments were all calculated by using this method.

4.2. Comparison of CPU/GPU time complexity

The results are shown in Figure 5, where the horizontal axis represents the training epochs and the vertical axis represents the training time in seconds. As observed, when training with both a CPU and GPU, we can clearly see the difference in the time consumption. The time spent when training with a GPU is significantly less than when training with a CPU. The processor was a 12th Gen $Intel^{(R)}$ $Core^{(TM)}$ i7-12700H 2.70 GHz, with a built-in RAM of 16.0 GB based on an x64 processor and an RTX 3060 GPU.

4.3. Comparison of F-measure result

The combination of BERT+BiLSTM+CRF provides an additional layer that is not in the BiLSTM-CRF model, as it includes BERT for initialization of word embeddings. This extra layer of BERT initialization is undoubtedly an advantage over random initialization. BERT+BiLSTM+CRF performance was found to surpass that of BERT+CRF. First, BERT employs the transformer, which is based on self-attention. In this computation process, position information is weakened (relying solely on

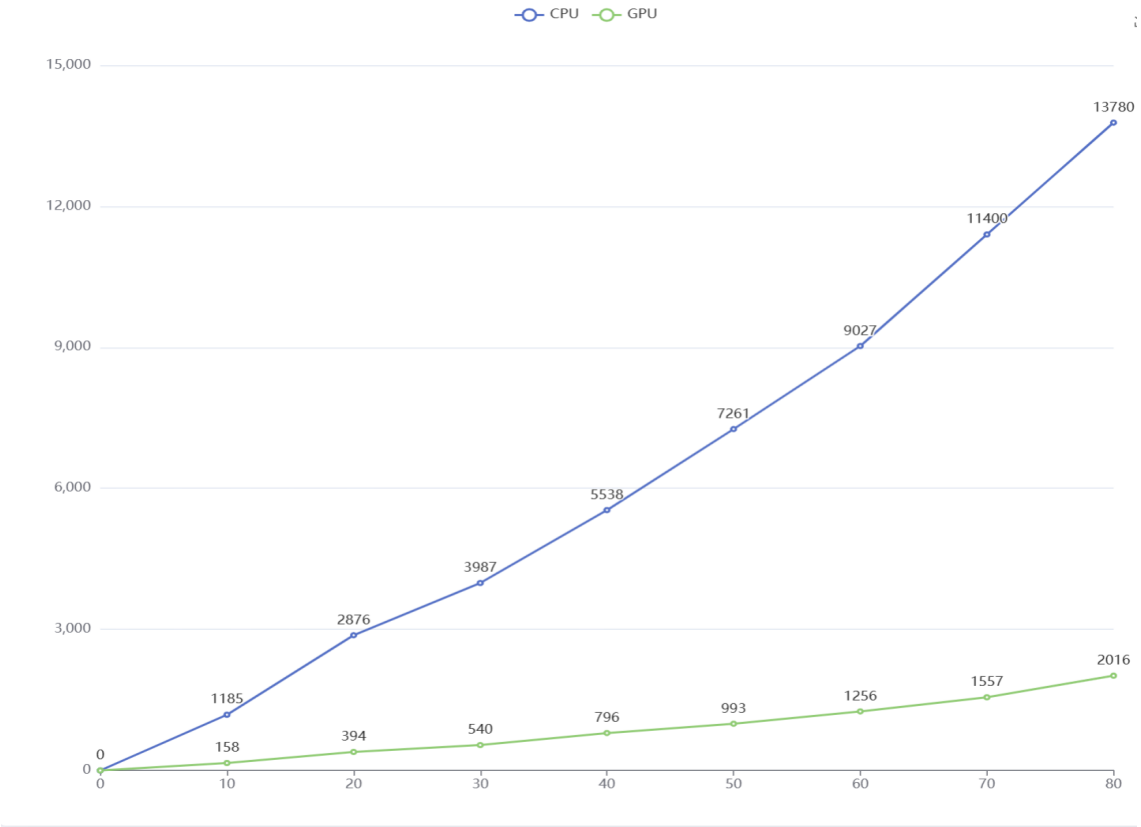

Figure 5. CPU/GPU time complexity.

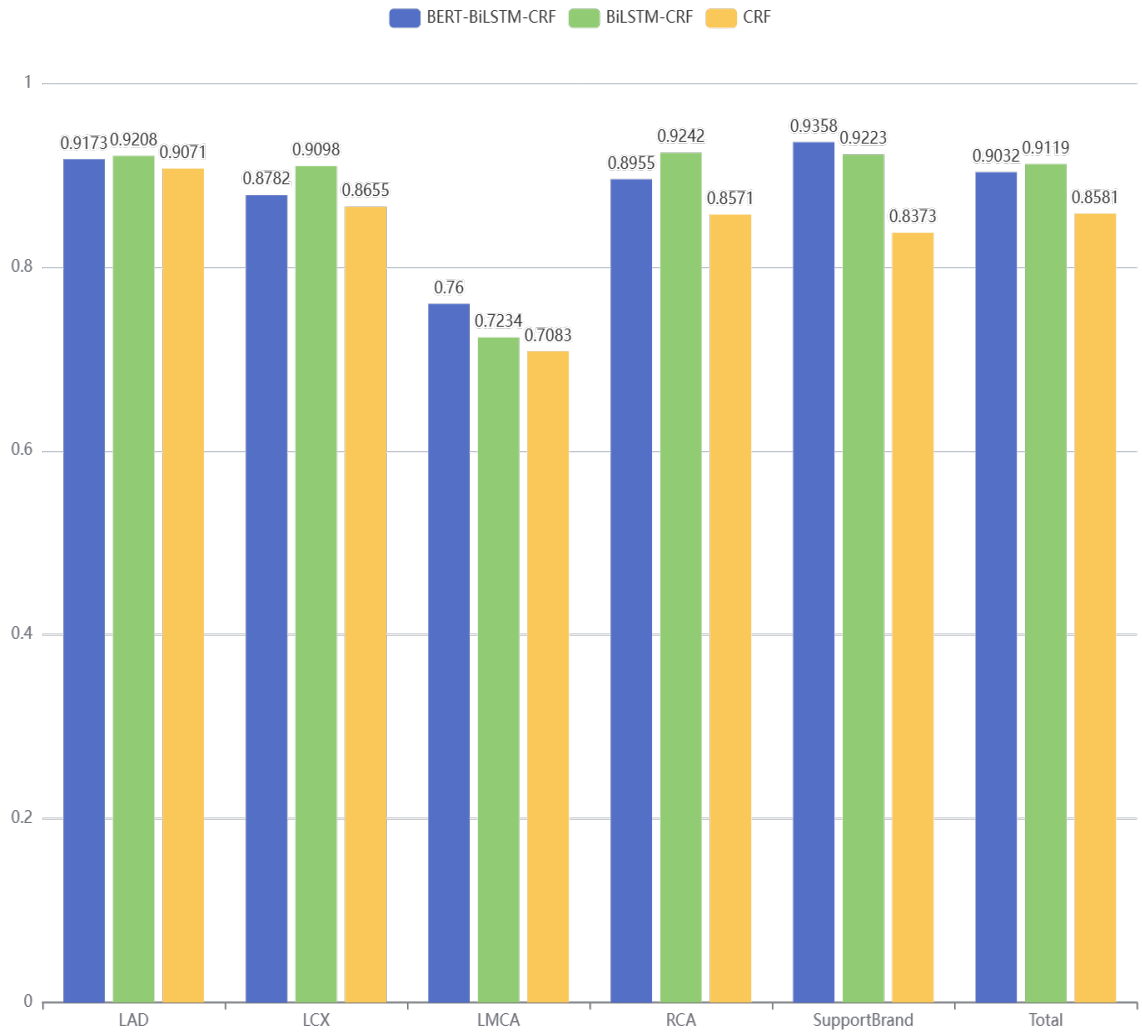

Figure 6. Comparison of F statistics results.

position embedding to inform the model about the position of input tokens). However, in the case of sequence labeling tasks, positional information is crucial, and even directional information is necessary. Therefore, we need to use LSTM to learn dependencies on the observed sequence, as well as to ultimately use a CRF model to learn the relationships between state sequences and obtain the answer. If the CRF is used directly, the model's learning capability on the observed sequence will decrease, leading to suboptimal performance.

The advantages of the BiLSTM-CRF model include the ability of the BiLSTM network to capture contextual information in the text, facilitating a better understanding of the context. When combined with the CRF layer, it can jointly model the entire sequence, taking into account dependencies between labels. Compared to some complex pre-trained models, BiLSTM-CRF is relatively simple, making training and understanding more intuitive. However, its drawback lies in BiLSTM's hyperfocus on capturing local information as it can, potentially fail to effectively capture global semantics. It requires a substantial amount of annotated data for supervised learning, and it may perform poorly in scenarios with scarce data for certain tasks.

On the other hand, the advantages of BiLSTM-BERT-CRF lie in BERT providing powerful contextual representations. The contextual information learned through pre-training contributes to a better understanding of the text. BERT excels on various NLP tasks, so BiLSTM-BERT-CRF might exhibit better generalization performance on specific tasks. Pre-trained models often achieved good results with relatively less annotated data. However, its disadvantages include the relatively large size of the BERT model, which demands significant computational resources and memory for training and inference. Pre-trained models may not be suitable for certain domain-specific tasks as they are trained on general corpora.

The results, as shown in Figure 6, present the F1 scores obtained by the three models for different labels. It is evident from this bar chart that, considering our dataset, the overall performance in terms of F1 score was the best for the BiLSTM-CRF model. The difference in F1 scores between the BiLSTM-CRF model and the BERT-BiLSTM-CRF model is not significant. However, when compared, the CRF model appears to have the least effective performance among the three models.

The LMCA label may not have performed as well due to the limited dataset from the EMRs that we used. This could have resulted in suboptimal training results for this specific label. However, it is worth noting that this label may not be extensively utilized in the medical context of focus here, so the results are within an acceptable range for our purposes.

Furthermore, from the figure, we can observe that the overall F1 score achieved by training with BiLSTM-CRF reached 91.19%. This result meets our expectations and is highly promising for application in our subsequent medical tasks.

5. Conclusions

In this study, we focused on the extraction of clinical information related to coronary angiography from Chinese EMRs. We employed three different models for clinical NER in coronary angiography. Our experiments demonstrate that deep learning methods can automatically extract clinical entities from EMRs. Among the three models, the BiLSTM-CRF model performed the best on our dataset. EMRs contain rich diagnostic information, which holds significant potential for our medical applications. We utilized machine learning methods to automatically retrieve coronary artery intervention

information from a large dataset of electronic health records of Chinese individuals. We have automated the extraction of coronary artery anatomy and lesion information, and calculated the Gensini score automatically. This approach facilitates clinical data research and provides a new technological tool for medical decision support.

In our future work, we address the limitations of our current models, which were mainly tailored for training on Chinese data. We recognize that achieving optimal results in English may pose challenges. Therefore, we will strive to obtain diverse surgical information from multiple hospitals and train our model accordingly. Our aim is to alleviate the burden on medical professionals. Furthermore, delving into the modification of medical entities and understanding entity relationships is a crucial objective for our next steps. Building upon our current work, we plan to further study similar medical records and conduct predictive analysis on the obtained results.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence tools in the creation of this article.

Acknowledgment

This research was supported by "Pioneer" and "Leading Goose" R&D Program of Zhejiang under Grants 2024C03096.

Author contributions

Dr. Feng Li proposed the scheme and designed the framework. Mingfeng Jiang, Yi Chen, Feng Chen and Wei Nie conducted the algorithms and verified the performances of the proposed model. Dr. Hongzeng Xu provided the medical guidance. Dr. Li Wang provided the methodology guidance and improved the presentation of this paper.

Conflict of interest

The authors declare that there is no conflict of interest.

References

1. T. Wang, P. Xuan, Z. Liu, T. Zhang, Assistant diagnosis with Chinese electronic medical records based on CNN and BiLSTM with phrase-level and word-level attentions, *BMC Bioinf.*, **21** (2020). https://doi.org/10.1186/s12859-020-03554-x
2. J. Tsai, G. Bond, A comparison of electronic records to paper records in mental health centers, *Int. J. Qual. Health Care*, **20** (2008),136–143. https://doi.org/10.1093/intqhc/mzm064
3. Y. Hu, Research on the information diagnostic technology based on medical information, University of Electronic Science and Technology of China, 2015.
4. Z. Obermeyer, E. J. Emanuel, Predicting the future—big data, machine learning, and clinical medicine, *N. Engl. J. Med.*, **375** (2016), 1216–1219. https://doi.org/10.1056/NEJMp1606181

5. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, **521** (2015), 436–444. https://doi.org/10.1038/nature14539

6. J. Yang, Y. Guan, B. He, C. Qu, Q. Yu, Y. Liu, et al., Corpus construction for named entities and entity relations on chinese electronic medical records, *J. Softw.*, **27** (2016), 2725–2746.

7. L. R. Rabiner, A tutorial on hidden Markov models and selected applications in speech recognition, *Proc. IEEE*, **77** (1989), 257–286. https://doi.org/10.1109/5.18626

8. A. Roberts, R. Gaizauskas, M. Hepple, Extracting clinical relationships from patient narratives, in *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, (2008), 10–18. https://doi.org/10.3115/1572306.1572309

9. J. Lafferty, A. McCallum, F. C. N. Pereira, Conditional random fields: Probabilistic models for segmenting and labeling sequence data, in *Proceedings of the 18th International Conference on Machine Learning 2001 (ICML 2001)*, (2001), 282–289. https://repository.upenn.edu/handle/20.500.14332/6188

10. S. Hochreiter , J. Schmidhuber , Long short-term memory, *Neural Comput.*, **9** (1997), 1735–1780.

11. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, preprint, arXiv: 1810.048052018.

12. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, et al., Deep contextualized word representations, *Assoc. Comput. Linguist.*, **1** (2018), 2227–2237. https://doi.org/10.18653/v1/N18-1202

13. T. Younga, D. Hazarikab, S. Poriac, E. Cambriad, Recent trends in deep learning based natural language processing, *IEEE Comput. Intell. Mag.*, **13** (2018), 55–75. https://doi.org/10.1109/MCI.2018.2840738

14. L. Ouyang, Y. Tian, H. Tang, B. Zhang, Chinese named entity recognition based on B-LSTM neural network with additional features, in *International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage*, (2017), 269–279. https://doi.org/10.1007/978-3-319-72389-1_22

15. Y. Xiang, Chinese named entity recognition with character-word mixed embedding, in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, (2017), 2055–2058.

16. H. Yang, H. Gao, Toward sustainable virtualized healthcare: Extracting medical entities from Chinese online health consultations using deep neural networks, *Sustainability*, **10** (2018), 3292. https://doi.org/10.3390/su10093292

17. W. Zhang, S. Jiang, S. Zhao, K. Hou, Y. Liu, L. Zhang, A BERT-BiLSTM-CRF model for Chinese electronic medical records named entity recognition, in *2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA)*, (2019), 166–169. https://doi.org/10.1109/ICICTA49267.2019.00043

18. X. Zhang, Y. Zhang, Q. Zhang, Y. Ren, T. Qiu, J. Ma, et al., Extracting comprehensive clinical information for breast cancer using deep learning methods, *Int. J. Med. Inf.*, **132** (2019), 103985.

19. L. Li, L. Jin, Y. Jiang, D. Huang, Recognizing biomedical named entities based on the sentence vector/twin word embeddings conditioned bidirectional LSTM, in *Chinese Computational Lin-*

guistics and Natural Language Processing Based on Naturally Annotated Big Data, (2016), 165–176. https://doi.org/10.1007/978-3-319-47674-2_15

20. M. Habibi, L. Weber, M. Neves, D. L. Wiegandt, U. Leser, Deep learning with word embeddings improves biomedical named entity recognition, *Bioinformatics*, **33** (2017), i37–i48. https://doi.org/10.1093/bioinformatics/btx228
21. J. P. C. Chiu, E. Nichols, Named entity recognition with bidirectional LSTM-CNNs, *Trans. Assoc. Comput. Linguist.*, **4** (2016), 357–370. https://doi.org/10.1162/tacl_a_00104
22. L. Li, Y. Guo, Biomedical named entity recognition with CNN-BLSTM-CRF, *J. Chin. Inf. Newsp.*, (2018), 116–122.
23. D. S. Sachan, P. Xie, M. Sachan, P. Xing, Effective use of bidirectional language modeling for transfer learning in biomedical named entity recognition, in *Proceedings of the 3rd Machine Learning for Healthcare Conference*, (2018), 383–402.
24. E. F. Tjong K. Sang, J. Veenstra, in *Proceedings of the Ninth Conference on European Chapter of the Association for Computational Linguistics*, (1999), 173–179. https://doi.org/10.3115/977035.977059
25. X. Dong, S. Chowdhury, L. Qian, Y. Guan, J. Yang, Q. Yu, Transfer bi-directional LSTM rnn for named entity recognition in chinese electronic medical records, in *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, (2017), 12–15. https://doi.org/10.1109/HealthCom.2017.8210840

AIMS Press