



Research article

Chinese medical event detection based on event frequency distribution ratio and document consistency

Ruirui Han, Zhichang Zhang*, Hao Wei and Deyue Yin

College of Computer Science and Engineering, Northwest Normal University, 967 Anning East Road, Lanzhou 730070, China

* **Correspondence:** Email: zzc@nwnu.edu.cn; Tel: +8613038769329.

Abstract: Structured information especially medical events extracted from electronic medical records has extremely practical application value and play a basic role in various intelligent diagnosis and treatment systems. Fine-grained Chinese medical event detection is crucial in the process of structuring Chinese Electronic Medical Record (EMR). The current methods for detecting fine-grained Chinese medical events primarily rely on statistical machine learning and deep learning. However, they have two shortcomings: 1) they neglect to take into account the distribution characteristics of these fine-grained medical events. 2) they overlook the consistency in the distribution of medical events within each individual document. Therefore, this paper presents a fine-grained Chinese medical event detection method, which is based on event frequency distribution ratio and document consistency. To start with, a significant number of Chinese EMR texts are used to adapt the Chinese pre-training model BERT to the domain. Second, based on the fundamental features, the Event Frequency - Event Distribution Ratio (EF-DR) is devised to select distinct event information as supplementary features, taking into account the distribution of events within the EMR. Finally, using EMR document consistency within the model improves the outcome of event detection. Our experiments demonstrate that the proposed method significantly outperforms the baseline model.

Keywords: Chinese electronic medical record; medical event detection; event frequency - event distribution ratio; document consistency

1. Introduction

EMR, or electronic medical record, is a critical source of clinical information generated by doctors and patients within medical facilities. It encompasses a wealth of medical knowledge that is closely tied to the health of the patient [1]. Structured medical information extracted from unstructured or semi-structured EMR text can be used for several purposes, such as aiding in clinical diagnoses and

treatment, monitoring and controlling diseases, evaluating treatment efficacy and conducting clinical research [2–4]. Therefore, the field of natural language processing and medical informatics has been focused on the structurization of EMRs for many years, making it a highly sought-after area of research. One crucial aspect of EMR structurization is medical event detection, which has garnered significant attention and interest. A number of recent studies proved that it has important practical application significance for the text structure of EMRs. In the work of people like Geva et al. [5], they explored the incidence of adverse drug events in children with pulmonary hypertension (PH) by comparing different sources of EMRs, which is of great significance for improving the treatment of children with pulmonary hypertension. Kulshrestha et al. [6] effectively predicted severe chest injuries by EMR data with NLP technology, which improves accuracy and helps the treatment and diagnosis of chest injuries. Ahuja et al. [2] also predicted the disease and activity level of multiple sclerosis (MS) based on EMRs, which can help doctors identify the possibility of disease progression so early as possible to take earlier intervention measures to improve the treatment effect. In recent years, the topic of automatically detecting medical events has captured wide attention from the research community, attracting numerous researchers and yielding a substantial amount of research output.

A medical event refers to any occurrences or circumstances related to a patient’s clinical timeline, providing insight into their physical health and medical process. It encompasses a wide range of information, including the patient’s medical history, symptoms, diagnostic evaluations, treatments and other related aspects of their care. Medical event detection is commonly recognized as a task of sequence labeling, which entails not only identifying the boundaries of medical events, but also categorizing their types. As depicted in Figure 1, given a text “患者自述腹部胀痛有所缓解 (The patient reports some relief in abdominal discomfort)”, the task of medical event detection requires identifying the “Evidence” type event “自述” and the “Problem” type event “腹部胀痛” in the text. Furthermore, the detection process must also mark the boundaries and type labels of the events.

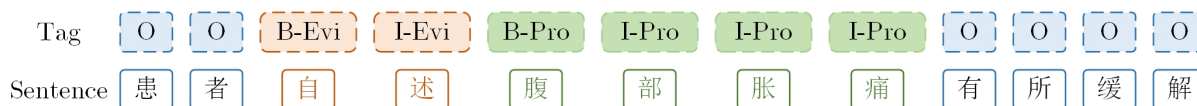


Figure 1. Illustration of medical event detection in Chinese. This paper employs the BIO annotation method, where “B-XXX” signifies the start of an event, “I-XXX” indicates the middle or conclusion of the event, “O” represents non-events and “XXX” represents different types of events. Specifically, “Evi” represents events of type “Evidence” and “Pro” stands for events of type “Problem”.

At present, the research on medical event monitoring of Chinese EMRs is still relatively weak and mainly divided into three categories. One class is rule-based method. This method relies on the experience and knowledge of domain experts to summarize relevant rules, and uses the corresponding template as a classifier to extract medical events. The other category is the method based on shallow machine learning. This method establishes a model to extract medical events through feature statistics. The last one is the most commonly used—deep learning-based method at present. This method obtains a more comprehensive contextual semantic representation through a deep neural network model, and its event extraction effect is better. Nevertheless, there are three shortcomings in the current research on Chinese medical event detection: 1) The BERT model used is trained on general domain data and

lacks prior medical knowledge when applied to medical event detection tasks, limiting its overall effectiveness; 2) The BERT model does not account for the relationship between masked words during training the mask language model, and only considers basic word and syntactic features, without taking into account the distributional characteristics of medical event words in EMRs; 3) Medical events in EMRs often follow document-consistent distribution patterns, meaning that there are specific distribution rules between different events, but existing studies only detect events at the sentence level and do not use document-level features. Additionally, while current research on medical event detection primarily focuses on English EMRs, research on Chinese EMRs is limited. In light of this, this paper puts forth a novel detection method that leverages the event frequency ratio and document consistency to address the limitations of current methods in detecting medical events in Chinese EMR. The results of experiments conducted on the Chinese Clinical Event Detection Corpus (Chinese CED) [7] demonstrate that the proposed method outperforms existing methods in the task of medical event detection.

The primary contributions of this research paper are as follows: 1) A significant number of Chinese medical texts were employed to fine-tune the pre-trained BERT model to the domain-specific context. 2) The EF-DR was devised based on basic features, and the distribution features of distinct event words were chosen as additional features of the corpus. 3) Document consistency distribution features were used to improve further the overall outcome. The experimental results show that the proposed method is significantly improved compared with the baseline method, thus highlighting the superiority of the proposed approach.

2. Related work

The task of evaluating the Informatics for Integrating Biology and the Bedside (i2b2) in 2012 involved detecting medical events such as patient issues, tests and treatments within a corpus of 310 English hospital discharge abstracts that had been manually annotated. Since 2015, the international semantic evaluation conference SemEval has held a technical evaluation competition for extracting clinical medical events for three consecutive years. The competition has made use of a publicly available evaluation dataset, which includes the THYME-2015 [8], THYME-2016 [9], THYME-2017 corpora [10]. In 2020, the technical evaluation of the China Conference on Knowledge Graph and Semantic Computing (CCKS) established the task of “Medical Entity and Event Extraction”. The main entity for this task was the Chinese EMR text data related to tumors. The task defined several attributes of tumors, such as tumor size and primary site, and required the identification and extraction of events and attributes, as well as the structuring of the Chinese EMR text. These evaluation tasks have significantly advanced the research in medical event detection.

This section provides a comprehensive overview and evaluation of the medical event detection methods presented in current research, which can be classified into three distinct categories.

1) Rule-based approach: This approach categorizes medical events based on their context. It employs various features such as context vocabulary, part of speech, tense of events, and an external knowledge base in the clinical medical field [11–13]. However, the performance of these methods is limited and they tend to perform poorly [14], making them inadequate for handling new types of data [15].

2) Statistical Machine Learning-Based Approach: These methods employ statistical machine learning techniques that rely on feature engineering, where SVM and CRF are the prominent methods

used [16]. However, a significant drawback of these methods is their strong reliance on manually labeled features, which requires a significant amount of time and financial resources, making it challenging to implement in practical applications. In the 2012 i2b2 review, nine out of the top ten systems used Conditional Random Fields (CRF) for event detection [17]. For classifying event attributes, the majority of systems employed Support Vector Machines (SVM) [18]. Four of the systems employed a hybrid approach, combining both statistical machine learning and rule-based methods [19–21].

3) Deep learning-based approach. These methods are favored by most researchers due to their lower dependence on external features and proven experimental results. A variety of neural network models have been introduced for the task of medical event detection, with the BiLSTM-CRF model being the most commonly used for early sequence labeling tasks [22]. Other models include Lattice-LSTM [23] and WC-LATM [24]. Ji et al. have applied the BiLSTM-CRF model to the recognition of named entities in Chinese EMR [25] and extraction of Chinese medical events [26]. Although the models mentioned above have shown impressive results in detecting medical events, they tend to overlook the significance of contextual information in semantic understanding. As a result, some researchers have developed models like BERT [27] and RoBERTa [28] that incorporate effective contextual information through pre-training. In 2020, Zhang et al. [29] used the RoBERTa-BiLSTM-CRF model for event extraction, while Dai et al. [30] used the entity vocabulary as a key feature in the input of the Chinese RoBERTa-wwm-ext-large pre-trained model, which ultimately earned them first place in the CCKS2020 evaluation Task 3. Yang et al. [31] in view of the lack of synergistic consideration of the global and local features of medical text information in existing models, proposed a hybrid neural network model based on CNN-BiLSTM-CRF (BCBC), which achieves the comprehensive extraction of the text's local and global features, and makes up for the lack of semantic capture due to the single model in traditional methods.

There have been limited studies on medical event detections in China, which primarily rely on sentence-level information without taking into account the distribution of event words within the text of EMRs and ignoring the inter-document feature information. In contrast, research on event detection in the news domain has been productive, with some researchers in the field of Automatic Content Extraction (ACE) news events suggesting improvements in event detection outcomes through leveraging document consistency across multiple documents discussing the same subject matter. Document consistency in a news corpus refers to the consistent distribution of event types within news articles that share the same topic. This means that if a particular type of event is present in a document, other related events are often also present. As a result, some researchers use the relationship between document-level events to enhance event detection in general domains and use information on the distribution of document-level events to aid in reasoning, thereby enhancing the accuracy of sentence-level event detection [8, 32, 33]. Although the language used in the medical field differs from that in journalism, there is still a correlation between events. In 2020, Wang et al. [34] used the document consistency feature in medical texts to enhance event distribution. They reclassified trigger words with low classification probability using a Support Vector Machine (SVM) and saw an improvement of 1.34% in the *F1* score. This shows that by incorporating document consistency, the information between sentences can be better captured, thus leading to improved experimental results.

3. Methods

In this study, we consider medical event detection as a sequence labeling task, and use the BERT-BiLSTM-CRF model, which has demonstrated exceptional performance in prior research, to annotate sequences in Chinese EMRs. The proposed method is comprised of three components: basic features, extended features, and document consistency, as illustrated in Figure 2. In the section on basic features, this work employs a vast amount of EMR data to pre-train a Chinese BERT model [35], with the aim of achieving domain adaptation. The output from the final hidden layer of the model serves as the basic features. For the extended feature component, the EF-DR is designed.

To begin, the medical event word dictionary in the EMR is processed through the natural language processing tool LTP [36] for segmentation, which allows us to identify the part of speech (POS) of each word. Subsequently, we use a coding tool to encode each word and its corresponding POS. Lastly, we use the EF-DR to assign weights to the encoded words and combine the part of speech codes to form an extended feature. In the Document Consistency Module, the basic and extended features are combined and served as input. The initial sequence marking result is obtained through decoding, and if the probability of the optimal sequence path reaches the set threshold value, it is used as the final result. If not, the Document Consistency features are added and decoding is performed again until the probability of the optimal sequence path meets the threshold value.

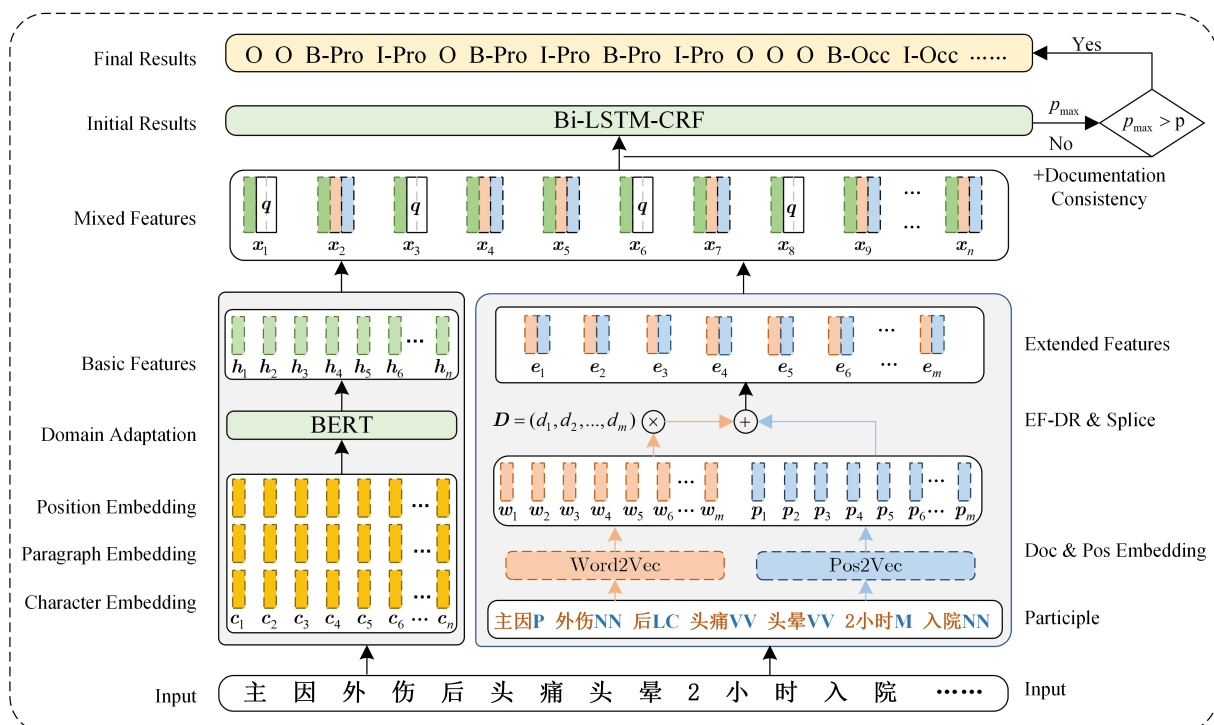


Figure 2. Illustration of Chinese medical event detection method based on event frequency distribution ratio and document consistency.

3.1. Basic features

In light of the fact that the amount of training data that has been manually labeled is often limited and plagued by substantial noise, the presented method aims to improve the quality of training data in a task-independent manner. The approach begins by carrying out synonym substitution, but only on the non-event parts to prevent any adverse effects on the event label. Afterwards, the revised text is generated by randomly rearranging the sentences in the original text. In sequence labeling tasks, the local labels are highly dependent on the context, and random rearrangement of words can result in significant boundary issues. Therefore, this method only performs random shuffling at the sentence level to mitigate these problems.

This work employs the Chinese BERT pre-training model, which has been trained on a massive corpus that includes Wikipedia and various books. Despite the fact that the BERT model has yielded impressive results in downstream tasks in the general domain, its lack of enough prior knowledge in the medical domain hampers its performance in medical event detection. To address this issue, the proposed approach adapts the BERT model to the medical domain by pre-training it on a large corpus of Chinese EMR texts [37]. This pre-training process allows the BERT model to acquire more prior knowledge in the medical field, thereby improving its ability to capture rich semantic information in obscure Chinese medical terms, leading to a more effective medical event detection.

The primary feature involves encoding the enhanced EHR text to produce the sequence of word codes $C = (c_1, c_2, \dots, c_n)$. Each word code c_i is a combination of three different types of embeddings, including BERT encoding with field adaptation, where the output from the last hidden layer is used. This forms the basic feature $H = (h_1, h_2, \dots, h_n)$, as shown in Eq (3.1).

$$H = \text{BERT}(c_1, c_2, \dots, c_n) = (h_1, h_2, \dots, h_n) \quad (3.1)$$

3.2. Event frequency - event distribution ratio

We have designed the EF-DR weighting scheme in order to fully use the word and POS information features, in addition to the basic features. The EF-DR leverages class distribution information of events in the training data to enhance the vector representation of words and parts of speech. This method gives higher importance to words with high frequency and unique classes observed in the training data, thus resulting in a more improved representation.

The EF-DR is a scheme of weighting that takes into account both the significance of an event word within its category (Event Frequency, EF) and the distribution of the category that the event word belongs to across all categories (Distribution Ratio, DR). For every event word in each EMR, the EF-DR is calculated. To simplify the calculation, the value of the EF-DR for non-event words is set to 1. The EF is calculated by determining the proportion of event words in a particular category, while the DR is calculated by finding the ratio of the occurrences of event word a in category b . The final EF-DR is obtained by multiplying these two measures, as showed in Eqs (3.2)–(3.4).

$$EF_a = \frac{a_b}{N_b} \quad (3.2)$$

$$DR_a = \frac{a_b}{M_a} \quad (3.3)$$

$$EF - DR_a = \frac{a_b}{N_b} \times \frac{a_b}{M_a} = \frac{a_b^2}{N_b \times M_a} \quad (3.4)$$

where a_b is the number of occurrences of event word a in category b , N_b is the total number of event words b in category b , and M_a is the total number of occurrences of event word a in all categories.

Record the EF-DR as $D = (d_1, d_2, \dots, d_m)$, where non-event words are given a default value of 1 for the EF-DR. The EF-DR assigns higher weights to event words that are unique to a particular category and have a high frequency, and lower weights to event words that have low frequency within a category or appear in multiple categories.

To mitigate the adverse effects of improper word segmentation, this study builds a medical event word dictionary using 4000 Chinese medical named entities. The collected medical event word dictionary is then integrated into the user dictionary of the natural language processing tool LTP, which is used to segment the corpus and perform POS tagging. Doc2vec [38] (see the original literature for details) and Pos2vec tool were respectively trained and used to convert the segmented corpus into event word code $W = (w_1, w_2, \dots, w_m)$ and part-of-speech code $P = (p_1, p_2, \dots, p_m)$. The final extended feature $E = (e_1, e_2, \dots, e_m)$ is obtained by multiplying the EF-DR D and concatenating the part-of-speech feature P (As shown in Eqs (3.5) and (3.6)).

$$e_i = [d_i \times w_i p_i] \quad (3.5)$$

$$E = ([d_1 \times w_1; p_1], [d_2 \times w_2; p_2], \dots, [d_m \times w_m; p_m]) = (e_1, e_2, \dots, e_m) \quad (3.6)$$

The final feature, X , is a combination of the basic and extended features, represented as $X = (x_1, x_2, \dots, x_m)$. The adopted strategy is to concatenate the extended feature to the basic feature representation of the last word in the event word, with the rest of the dimensions filled using the padding (PAD) vector q . This is done to maintain consistent dimensions and ease subsequent operations. The final input feature X is depicted in Eq (3.7).

$$X = ([h_1; q], [h_2; e_1], \dots, [h_n; e_m]) = (x_1, x_2, \dots, x_n) \quad (3.7)$$

3.3. Documentation consistency

This work focuses on two types of EMR texts: the discharge summary and the first course record. These texts are typically written in a specific order, starting with the chief complaint and then covering the patient's past, physical examination, auxiliary examination, diagnosis, and treatment. Using "Exam" event as an example, the results showed that there is a higher probability of "Problem" events before the "Exam" event, while the probability of "Treatment" events, "Evidence" events, and "Aspectual" events is higher after the "Exam" event. The distribution of different event types before and after the "Exam" event in EMR is shown in Figure 3. To further improve the detection results, this paper leverages the document consistency feature of EMR by increasing the frequency distribution ratio of events.

We denote the sequence of label types as $L = l_1, l_2, l_3, \dots, l_{15}$ (7 types of events, using BIO tags). The feature X input BiLSTM to obtain the probability value R_{wl} of each word w under each label l , the CRF part will provide the transition probabilities T_L , the probability of recording a single path is

p . Enter the sequence $c = (c_1, c_2, \dots, c_n)$. In the path $(l_1^1, l_1^2, \dots, l_1^n)$ (That is, each word w all correspond to the label type l_1). The probability is shown in Eqs (3.8)–(3.10).

$$R_{wL} = R_{w_1 l_1} + R_{w_2 l_1} + \dots + R_{w_n l_1} \quad (3.8)$$

$$T_L = T_{l_1, l_1}^1 + T_{l_1, l_2}^2 + \dots + T_{l_1, l_2}^{n-1} \quad (3.9)$$

$$p = R_{wL} + T_L \quad (3.10)$$

Among, R_{w_i, l_i} representation word c_i in the label type l_i the probability value under, T_{l_1, l_1}^i express c_i corresponding label type l_1 To c_{i+1} Corresponding label type l_1 the transition probability under. From this, we can obtain the probability value of each input sequence under all paths, taking the probability value of the optimal path p_{max} . If p_{max} Less than the set threshold 1 the full-text code information is added, and that document consistency features is use for re-labeling; If p_{max} Greater than the set threshold p the initial result is output as the final sequence annotation result.

4. Experiments

4.1. Dataset

Figure 4 shows an example of the data used in this paper. First, the EMR is annotated to obtain the annotated EMR [39], and then preprocessed to obtain a data format that conforms to the model input.

In this study, 2000 Chinese EMRs were randomly divided into three datasets: a training dataset (8:1:1 ratio), a validation dataset and a test dataset. The dataset encompasses seven types of events, including Problem, Exam, Treatment, Clinical Department, Evidence, Occurrence and Aspectual, as indicated in Table 1. The distribution statistics of other event types appearing before and after the inspection event type are displayed in Figure 3, for each EMR.

Table 1. Detailed statistics of Chinese CED corpus.

Type	Training dataset	Test dataset	Validation dataset	total	Percentage (%)
Aspectual	2511	325	358	3194	2.74
Clinical department	125	19	16	160	0.14
Evidence	6978	851	929	929	7.52
Exam	19,925	2445	2602	24,927	21.4
Occurrence	3248	393	459	4100	3.52
Problem	49,884	6099	6768	62,751	53.87
Treatment	10,079	1145	1359	12,599	10.82

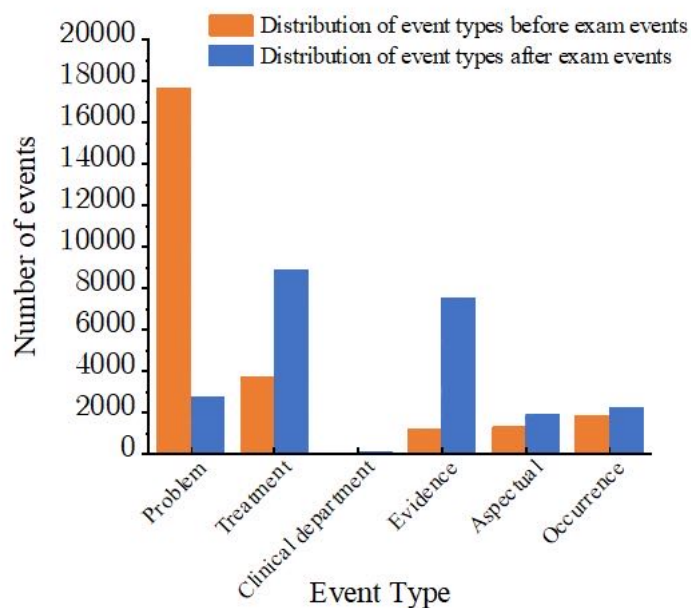


Figure 3. Distribution statistics of various types of events before and after “Exam” type events in EMR.

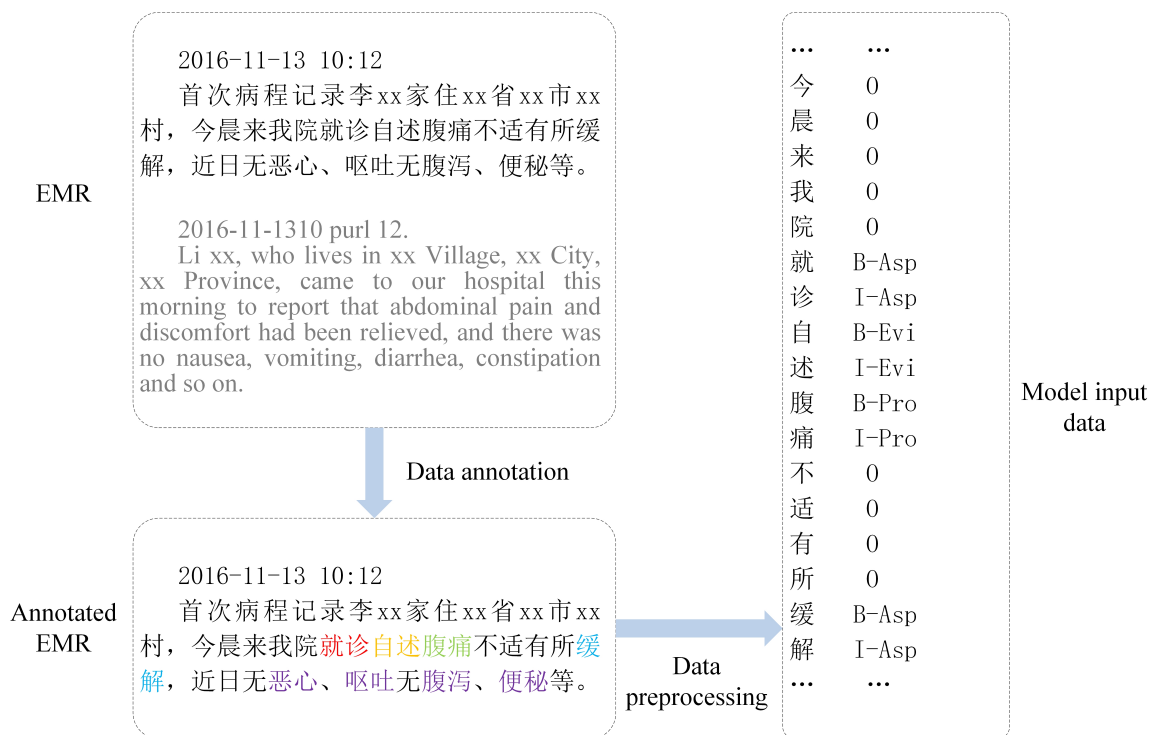


Figure 4. Example graph of model input data.

In EMR, there are variations in the way medical terms are written due to different writing habits among doctors. This, combined with the presence of abbreviations, can make follow-up work difficult. To address these issues, this paper performs basic data preprocessing. Irrelevant sentences are removed, redundant brackets in the EMR template are eliminated and misplaced punctuation marks are standardized. After preprocessing, a BERT-BiLSTM-CRF model is employed to detect medical events. In the baseline experiment, the output of the last hidden layer of a pre-trained Chinese BERT model, without any adaptation for the specific domain, is used as the fundamental feature. No additional event-related features are added. Based on the baseline, we selectively integrate the distribution features of various events using the EF-DR. Finally, for sequences whose optimal path probability calculated by CRF is below the threshold, we use the consistency feature of EMR documents, specifically the full-text coding information, to re-label them in order to enhance event detection performance.

4.2. Baseline model

In the experimental phase, three classic baseline models were used: CRF, BiLSTM-CRF and BERT-BiLSTM-CRF. Among them, the CRF model is a conventional sequence modeling technique that was commonly used before deep learning advanced significantly. Despite its simplicity, it requires manual feature creation. On the other hand, the BiLSTM-CRF model is an end-to-end deep learning approach that eliminates the need for manual feature engineering. The BiLSTM component of the model captures the context-based semantics of each word, and the CRF component differs from traditional CRF models. It retains the idea of a transition matrix from traditional CRF models but reimplements it in a deep learning context. The BERT-BiLSTM-CRF model differs from the BiLSTM-CRF model in the way it obtains word vectors. While the BiLSTM-CRF model uses a simple encoding layer learning to obtain word vectors, the word vectors in the BERT-BiLSTM-CRF model are generated during the training of downstream tasks. This model leverages the strengths of the Chinese pre-trained model BERT, namely its strong capability in acquiring dynamic word vectors. The rest of the layers in the BERT-BiLSTM-CRF model are identical to those in the BiLSTM-CRF model.

4.3. Domain adaptation

Table 2. Statistics of domain adaptation dataset.

Dataset	Total
EMR text	4000 (39.75 MB)
CCKS 2017	400 (0.32 MB)
CCKS 2018	1000 (0.80 MB)
CCKS 2019	1379 (1.1 MB)
CCKS 2020	1050 (0.84)

In the process of domain adaptation, we used data from a Class A hospital located in Gansu Province, China. The data covers 15 departments, including the cardiovascular, neurosurgery, pediatric obstetrics and gynecology, and infectious disease departments, and comprises a total of 4000 EMRs after removing sensitive information. The data was manually annotated with six types of medi-

cal named entities, including diseases, symptoms and abnormal examination results. Additionally, we also used the dataset from the CCKS EMR Named Entity Recognition task of the past four years, as depicted in Table 2.

During the domain adaptation process, no supervised training was incorporated and we used RoBERTa [29]. Specifically, we used the masked language model (MLM) task after removing the next sentence prediction (NSP) task. To preserve Chinese semantic information, the whole word mask (WWM) mechanism was employed. Furthermore, to enhance the training data, each sample was duplicated and re-masked to achieve dynamic masking. The domain adaptation was trained for 10 rounds with a batch size of 32. During the training process, a learning rate warm-up and linear decay strategy was employed. Specifically, the learning rate was increased linearly from 0 to $3e-5$ during the first 10% of the training steps, and then decreased linearly from $3e-5$ to 0 over the remaining 90% of the steps. The maximum sequence length was set to 510, and the optimization algorithm used was the Adam algorithm with a weight decay parameter of 0.01. For fine-tuning the event detection task, the maximum learning rate was set to $6e-5$, and the optimal parameters are presented in Table 3.

Table 3. Statistics of domain adaptation dataset.

Hyperparameters	Domain Adaptation (MLM)	Fine tuning (event detection)
Batch Size	32	16
Epochs	10	32
Peak learning Rate	$3e-5$	$6e-5$
Learning rate decline method	Linear	Linear
Learning Rate Warmup Ratio	0.05	0.1
Max Sequence Length	510	510
Weight Decay	0.01	0.01
CRF layer learning rate	N/A	$1e-3$

4.4. Evaluation index

In the model performance evaluation session, the model performance is evaluated using accuracy P , recall R and $F1$ value. The calculation formula is as shown in Eqs (4.1)–(4.3) (TP and FP are true and false positive cases respectively, TN and FN are true negative and false negative cases respectively).

$$P = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FP_i} \quad (4.1)$$

$$R = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n TP_i + \sum_{i=1}^n FN_i} \quad (4.2)$$

$$F1 = \frac{2 \times P \times R}{P + R} \quad (4.3)$$

4.5. Experimental results

In this paper, we carried out a series of experiments on the test dataset of the Chinese CED corpus. The results show that the method proposed in this paper outperforms all the baseline models, as evidenced by the data in Table 4. The experiments demonstrate that the accuracy of the proposed method is impressive, reaching 91.52% of P value and 92.43% of $F1$ value. In comparison to the BERT-BiLSTM-CRF model, the proposed method has improved the $F1$ value from 88.88% to 92.43%, a significant increase of 3.55%.

Table 4. Experimental results of overall performance.

Method	P (%)	R (%)	$F1$ (%)
CRF [7]	79.96	85.37	82.58
BiLSTM-CRF [7]	81.32	84.68	82.97
AdvBERT [7]	83.73	85.56	85.12
BERT-BiLSTM-Softmax	82.62	84.77	83.68
BERT-BiLSTM-CRF	87.30	90.52	88.88
Our Method	91.52	93.36	92.43

Table 5. Ablation experiment results.

Method	$F1$ (%)						P (%)	R (%)	$F1$ (%)
	Aspectual	Evidence	Exam	Occurrence	Problem	Treatment			
Our Method	93.16	90.76	91.34	95.85	93.52	86.55	91.52	93.36	92.43
-Document consistency	92.40	89.35	89.16	94.12	92.38	83.80	89.61	91.73	90.66
-EF-DR	91.81	89.86	89.87	95.12	91.55	82.43	88.98	91.59	90.27
-Extended features	91.11	87.75	89.25	94.87	91.33	83.24	88.96	90.94	89.94
-Domain adaptation	91.88	88.23	89.23	94.65	90.76	78.27	87.51	90.81	89.13
-Data enhancement	87.80	88.57	89.08	94.65	90.11	80.33	87.30	90.52	88.88

The results of the ablation experiments are shown in Table 5. As can be seen from Table 5, When the added data is enhanced, the overall $F1$ value is 0.25% higher than the baseline model. Among them, the effect of “Aspectual” type events detection has been significantly improved, However, $F1$ value decreased for “Treatment” type events. The possible reason is that the therapeutic event is more dependent on the context information contained in the preceding and following sentences. When the sentences are randomly shuffled, the context will be destroyed, which will lead to errors in the detection of therapeutic events. When the BERT model is adapted to the domain, the accuracy is improved by 1.45%. The $F1$ value increased by 0.81%. At this time, the detection effect of “Treatment” type events is not only restored but also greatly improved, indicating that domain adaptation enables the model to obtain more prior knowledge of the medical field. After adding lexical and part-of-speech information, the value increased by 0.32%. The addition of event frequency distribution ratio is helpful to describe the detection effect of “Evidence” type events, “Aspectual” type events and “Treatment” type events.

The $F1$ value increased by 0.4%. The use of document consistency feature makes the detection effect of “Problem”, “Evidence” and “Occurrence” type events significantly improved, and improved by 1.77% compared with the previous $F1$ value, which further proves the superiority of the method in this paper.

Compared to the baseline model, the proposed method has seen an improvement in performance for the following reasons: Firstly, the Chinese BERT pre-training model has been adapted to the domain through data enhancement, reducing the impact of factors such as limited data and high levels of noise. This allows the model to acquire more prior knowledge in the medical field. Secondly, to better use vocabulary, POS information, and event distribution features, we have designed the EF-DR weighting scheme. This not only considers the distribution of different events in EMR texts, but also allows full use of lexical and POS features. Finally, to further supplement document level features that are difficult to extract at the sentence level in event detection, we have employed the document consistency principle of EMR, resulting in improved event detection performance.

4.6. Error analysis

According to the event results predicted by our model, the following findings were discovered after conducting an error analysis, as illustrated in Table 6. The first issue identified was a nested event detection error. Nested events are a unique type of event that contain one or more other events within them. For instance, the event “心肺腹及神经系统查体” in Sentence 1 belongs to the “Exam” type event. The event “神经系统查体” can also be considered a separate “Exam” type event. From a human perspective, the examination event of “心肺腹查体” can be understood as “心肺腹查体” and “神经系统查体”. However, from the model’s perspective, it does not fully comprehend the deeper meaning of the sentence or the relationships between the objects connected by conjunctions such as “及”, “和”, etc.

The second issue is that the predicted event boundary does not match the event boundary labeled in the data, but it aligns with the medical event labeling specification. It is apparent that there are errors in the manually labeled data, yet the actual performance of the model in this study surpasses the performance indicated by the evaluation metrics. As stated in Sentence 3, two instances of “反复腹胀” were mistakenly labeled as a “Problem” type event, but the model accurately identifies the “Aspectual” type event “反复” and the “Problem” type event “腹胀”. Additionally, the “Occurrence” type event, which was missed in the manual tagging, is also mentioned in sentence four. Furthermore, the model in this study correctly recognizes the unlabeled events in the data, as evidenced by its ability to identify multiple “Treatment” type events and “Exam” type events in Sentences 5 and 6. This demonstrates that the model’s learning capabilities extend beyond what is shown by the evaluation metrics.

Through the analysis, the superiority of the method in this paper is further proved. Compared with the baseline model, in this paper, it enables the pre-trained language model BERT to obtain more background knowledge in the medical field through domain adaptation, which helps the model understand and deal with obscure professional expressions in EMRs. The event frequency distribution ratio designed in this paper has a positive impact on the final event detection accuracy through statistical methods: quantifying and expressing the event distribution information in the EMR, and through regulating weights of the event information. Combined with the characteristics of the sequence labeling task, and incorporated the consistency of EMRs, the global information of this model is further enriched, which helps the model to detect events more accurately.

Table 6. Error analysis.**Type 1: Nested event detection error.**

句子1: 查体: bp: 140/95 mmHg, 心肺腹及神经系统查体未见明显异常。

Sentence 1: Examination: bp: 140/95 mmHg, no significant abnormalities on cardiopulmonary and abdominal and neurological examination.

True label: 'Exam': '心肺腹及神经系统查体', 'bp', '查体'

True label: 'Exam': 'cardiopulmonary, abdominal and neurological examination', 'bp', 'Examination'

Predict label: 'Exam': 'bp', '查体', '神经系统查体', '心肺腹'

Predict label: 'Exam': 'bp', 'Examination', 'neurological examination', 'cardiopulmonary, abdominal'

句子2: 双肺未闻及病理性呼吸音及干湿性。

Sentence 2: No pathological breath sounds and dry moist sounds were heard in both lungs.

True label: 'Problem': '双肺未闻及病理性呼吸音及干湿性音'

True label: 'Problem': 'No pathological breath sounds and dry moist sounds were heard in both lungs.'

Predict label: 'Problem': '干湿性音', '双肺未闻及病理性呼吸音'

Predict label: 'Problem': 'dry moist sounds', 'No pathological breath sounds were heard in both lungs'

Type 2: Predictions are inconsistent with the true label but correct.

句子3: 主因反复腹胀, 腹痛不适半年余。

Sentence 3: The main cause was recurrent abdominal distension and abdominal pain and discomfort for more than six months.

True label: 'Problem': '反复腹胀', '腹痛'

True label: 'Problem': 'recurrent abdominal distension', 'abdominal pain'

Predict label: 'Problem': '腹胀', '腹', 'Aspectual': '反复'

Predict label: 'Problem': 'abdominal distension', 'abdominal', 'Aspectual': 'recurrent'

句子4: 患者诉于入院前5天与他人争吵后被他人殴打左前臂。

Sentence 4: The patient complained of being beaten on the left forearm by another person after an argument with another person 5 days before admission.

True label: 'Occurrence': '殴打', '入院'

True label: 'Occurrence': 'beaten', 'admission'

Predict label: 'Occurrence': '入院'

Predict label: 'Occurrence': 'admission'

Type 3: Unlabeled event detected.

句子5: 自动收缩患肢肌肉活动, 深呼吸运动, 多饮水, 积极预防并发症。

Sentence 5: Automatic contraction of muscle activity in the affected limb, deep breathing exercises, drinking more water, and active prevention of complications.

True label: Null.

Predict label: 'Treatment': '多饮水', '深呼吸运动', '自动收缩患肢肌肉活动'

Predict label: 'Treatment': 'drinking more water', 'deep breathing exercises', 'Automatic contraction of muscle activity in the affected limb'

句子6: 查体: bp: 130/70 mmHg。

Sentence 6: Examination: bp: 130/70 mmHg

True label: Null.

Predict label: 'Exam': 'bp'

Predict label: 'Exam': 'bp'

5. Conclusions

Aiming to address the issues with current medical event detection methods, this paper proposes a novel method for detecting Chinese medical events, using event EF-DR and document consistency. First, the training dataset of Chinese CED is enhanced, and the Chinese BERT pre-training model was fine-tuned through using a vast amount of Chinese EMR text data to capture more informative contextual features. Then, the EF-DR weighting scheme was developed to make better use of vocabu-

lary, part-of-speech information and event distribution features. Furthermore, a document consistency feature was introduced to complement the feature information between document-level events, further enhancing the event detection performance. As a result, the $F1$ value of this method reached 92.43%, which was 3.55% higher than that of the baseline model. This study enhanced the accuracy of structuring Chinese EMR and other medical texts, which has important practical significance for smart medical care. Including: 1) It can promote clinicians to make more accurate and reliable decisions in terms of treatment and medication, thereby improving medical quality. 2) It can help doctors find patient information faster, avoid repeated examinations and treatments and improve medical efficiency. 3) It can be used to track and analyze the development trend of large-scale diseases, and provide an important basis for the formulation and implementation of public health policies. 4) It can be used in medical research, providing large-scale disease data and promoting the development and progress of medical science.

Although the method in this paper has a good performance on the Chinese medical event detection task, there is still room for improvement. Through the observation and analysis of the experimental data and results, it is found that although the proportion of “Treatment” events is not low, the effect of its extraction is not good. Its accuracy rate and $F1$ value are lower than those of “Evidence”, “Aspectual” and “Occurrence” types of events whose proportion is not high. The follow-up plan of this paper will continue to study and propose a solution to this problem. In addition, the effect on the event type of the clinical department with less data is not outstanding, so another focus of our next work is to improve the recognition effect of the event in the clinical department under the condition of uneven data distribution. Furthermore, research on the temporal relationship of medical events in Chinese EMRs will be conducted in combination with temporal expressions, providing technical support for clinical applications such as disease monitoring and clinical assistant decision-making.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62163033), the Talent Innovation and Entrepreneurship Project of Lanzhou, China (No. 2021-RC-49), the Natural Science Foundation of Gansu Province, China (No. 21JR7RA781, No.21JR7RA116), the Major Research Project Incubation Program of Northwest Normal University, China (No. NWNLU-LKZD2021-06).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. J. H. Wang, *Research on Information Extraction Method of Chinese Electronic Medical Records*, Hunan University, 2016.
2. Y. Ahuja, N. Kim, L. Liang, T. Cai, K. Dahal, T. Seyok, et al., Leveraging electronic health records data to predict multiple sclerosis disease activity, *Ann. Clin. Transl. Neurol.*, **8** (2021), 800–810. <https://doi.org/10.1002/acn3.51324>

3. A. Geva, S. H. Abman, S. F. Manzi, D. D. Ivy, M. P. Mullen, J. Griffin, et al., Leveraging electronic health records data to predict multiple sclerosis disease activity, *J. Am. Med. Inf. Assoc.*, **27** (2020), 294–300. <https://doi.org/10.1093/jamia/ocz194>
4. Y. H. Su, C. P. Chao, L. C. Hung, S. F. Sung, P. J. Lee, A natural language processing approach to automated highlighting of new information in clinical notes, *Appl. Sci.*, **10** (2020), 2824. <https://doi.org/10.3390/app10082824>
5. A. Geva, S. H. Abman, S. F. Manzi, D. D. Ivy, M. P. Mullen, J. Griffin, et al., Adverse drug event rates in pediatric pulmonary hypertension: a comparison of real-world data sources, *J. Am. Med. Inf. Assoc.*, **27** (2020), 294–300. <https://doi.org/10.1093/jamia/ocz194>
6. S. Kulshrestha, D. Dligach, C. Joyce, M. S. Baker, R. Gonzalez, A. P. O'Rourke, et al., Prediction of severe chest injury using natural language processing from the electronic health record, *Injury*, **52** (2021), 205–212. <https://doi.org/10.1016/j.injury.2020.10.094>
7. Z. C. Zhang, M. Y. Zhang, T. Zhou, Y. L. Qiu, Pre-trained language model augmented adversarial training network for Chinese clinical event detection, *Math. Biosci. Eng.*, **17** (2020), 2825–2841. <https://doi.org/10.3934/mbe.2020157>
8. S. Bethard, L. Derczynski, G. Savova, J. Pustejovsky, M. Verhagen, Semeval-2015 task 6: Clinical tempeval, in *Proceedings of the 9th International Workshop on Semantic Evaluation*, (2015), 806–814.
9. S. Bethard, G. Savova, W. T. Chen, L. Derczynski, J. Pustejovsky, M. Verhagen, Semeval-2016 task 12: Clinical tempeval, in *Proceedings of the 10th International Workshop on Semantic Evaluation*, (2016), 1052–1062.
10. S. Bethard, G. Savova, M. Palmer, J. Pustejovsky, Semeval-2017 task 12: Clinical tempeval, *Proceedings of the 11th International Workshop on Semantic Evaluation*, (2017), 565–572.
11. *NegEx*, Available from: <http://code.google.com/p/negex/>.
12. P. Szolovits, Adding a medical lexicon to an English parser, in *AMIA Annual Symposium Proceedings*, (2003), 639.
13. T. C. Sibanda, *Was the Patient Cured?: Understanding Semantic Categories and their Relationship in Patient Records*, Massachusetts Institute of Technology, 2006.
14. Y. Sun, A. Nguyen, L. Sitbon, S. Geva, Rule-based approach for identifying assertions in clinical free-text data, in *Proceedings of 15th Australasian Document Computing Symposium*, (2010), 93–96.
15. A. L. Minard, A. L. Ligozat, A. B. Abacha, D. Bernhard, B. Cartoni, L. Deléger, et al., Hybrid methods for improving information access in clinical documents: concept, assertion, and relation identification, *J. Am. Med. Inf. Assoc.*, **18** (2011), 588–593. <https://doi.org/10.1136/amiajnl-2011-000154>
16. J. Lafferty, A. McCallum, F. 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)*, **8** (2001), 282–289. <https://doi.org/10.1109/ICIP.2012.6466940>
17. C. Cortes, V. Vladimir, Support-vector networks, *Mach. Learn.*, **20** (1995), 273–297. <https://doi.org/10.1023/A:1022627411411>

18. Y. Xu, Y. N. Wang, T. R. Liu, J. Tsujii, E. I. Chang, An end-to-end system to identify temporal relation in discharge summaries: 2012 i2b2 challenge, *J. Am. Med. Inf. Assoc.*, **20** (2013), 849–858. <https://doi.org/10.1136/amiajnl-2012-001607>
19. K. Roberts, B. Rink, S. M. Harabagiu, A flexible framework for recognizing events, temporal expressions, and temporal relations in clinical text, *J. Am. Med. Inf. Assoc.*, **20** (2013), 867–875. <https://doi.org/10.1136/amiajnl-2013-001619>
20. Y. K. Lin, H. Chen, R. A. Brown, MedTime: A temporal information extraction system for clinical narratives, *J. Biomed. Inf.*, **46** (2013), 20–28. <https://doi.org/10.1016/j.jbi.2013.07.012>
21. J. D’Souza, V. NgY, Classifying temporal relations in clinical data: a hybrid, knowledge-rich approach, *J. Biomed. Inf.*, **46** (2013), 29–39. <https://doi.org/10.1016/j.jbi.2013.08.003>
22. Z. H. Huang, W. Xu, K. Yu, Bidirectional LSTM-CRF models for sequence tagging, preprint, arXiv:math/1508.01991. <https://doi.org/10.48550/arXiv.1508.01991>
23. Y. Zhang, J. Yang, Chinese NER using lattice LSTM, preprint, arXiv:math/1805.02023. <https://doi.org/10.48550/arXiv.1805.02023>
24. W. Liu, T. G. Xu, Q. H. Xu, J. Y. Song, Y. R. Zu, An encoding strategy based word-character LSTM for Chinese NER, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, **1** (2019), 2379–2389. <https://doi.org/10.18653/v1/N19-1247>
25. B. Ji, R. Liu, S. S. Li, J. T. Tang, Q. Li, W. S. Xu, A BiLSTM-CRF method to Chinese electronic medical record named entity recognition, in *Proceedings of the 2018 International Conference on Algorithms*, **48** (2018), 1–6. <https://doi.org/10.1145/3302425.3302465>
26. B. Ji, R. Liu, S. S. Li, J. T. Tang, Q. Li, W. S. Xu, A joint extraction method for Chinese medical events, *Comput. Sci.*, **48** (2018), 287–293. <https://doi.org/10.48550/arXiv.1907.11692>
27. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, preprint, arXiv:math/1810.04805. <https://doi.org/10.48550/arXiv.1810.04805>
28. Y. H. Liu, M. Ott, N. Goyal, J. F. Du, M. Joshi, D. Q. Chen, et al., Roberta: A robustly optimized bert pretraining approach, preprint, arXiv:math/1907.11692. <https://doi.org/10.48550/arXiv.1907.11692>
29. X. N. Zhang, CCKS2020 medical event extraction based on named entity recognition. Available from: https://bj.bcebos.com/v1/conference/ccks2020/eval_paper/ccks2020_eval_paper_3_2_2.pdf.
30. S. T. Dai, Small sample medical event extraction based on pre-trained language model. Available from: https://bj.bcebos.com/v1/conference/ccks2020/eval_paper/ccks2020_eval_paper_3_2_1.pdf.
31. L. Y. Yang, J. Q. Li, Z. C. Zhu, X. M. Dong, F. Akhtar, Chinese medical event extraction based on hybrid neural network, in *2022 IEEE 46th Annual Computers*, (2022), 1420–1425. <https://doi.org/10.1109/COMPSAC54236.2022.00225>
32. L. Hou, P. Li, Q. Zhu, Study of event recognition based on CRFs and cross-event, *Comput. Eng.*, **38** (2012), 191–195.

33. J. B. Xia, C. Y. Fang, X. Zhang, A novel feature selection strategy for enhanced biomedical event extraction using the Turku system, *BioMed Res. Int.*, **2014** (2014), 1–12. <https://doi.org/10.1155/2014/205239>
34. C. Wang, P. Zhai, Y. Fang, Chinese medical event detection based on feature extension and document consistency, in *2020 5th International Conference on Automation, Control and Robotics Engineering*, (2020), 753–758. <https://doi.org/10.1109/CACRE50138.2020.9230246>
35. I. Turc, M. W. Chang, K. Lee, K. Toutanova, Well-read students learn better: On the importance of pre-training compact models, preprint, arXiv:math/1908.08962. <https://doi.org/10.48550/arXiv.1908.08962>
36. *LTP*, Available from: <http://ltp.ai/>.
37. S. Gururangan, A. Marasović, S. Swayamdipta, K. Lo, I. Beltagy, D. Downey, et al., Don't stop pretraining: Adapt language models to domains and tasks, preprint, arXiv:math/2004.10964. <https://doi.org/10.48550/arXiv.2004.10964>
38. Q. Le, T. Mikolov, Distributed representations of sentences and documents, in *International Conference on Machine Learning*, **32** (2014), 1188–1196.
39. M. Y. Zhang, *Research on Time Series Information Extraction Technology of Clinical Medical Events Based on Electronic Medical Records*, Northwest Normal University, 2021. <https://doi.org/10.27410/d.cnki.gxbfu.2021.001837>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)