International Journal of Robotics and Automation, Vol. 38, No. 10, 2023
An Open Access Paper

RELATION EXTRACTION METHOD OF CHINESE MEDICAL TEXT BASED ON RSIG-LSTM

Songpu Li,* Ruxin Gong,*** Xiaosheng Yu**,*** and Jiaheng Li****

Abstract

Purpose–long–short term memory (LSTM) is widely used in relation extraction. Tanh activation function in LSTM faces a vanishing gradient problem, which can hinder the transmission of knowledge and cause errors in the experimental results. Design/methodology/approach– in this paper, we propose a relation extraction method based on RSigELUS-LSTM (RSig-LSTM). Firstly, we use the bidirectional encoder representation from transformers (BERT) network model to embed word information. Secondly, we combine bidirectional RSig-LSTM with an attention mechanism to process features. We use the Softmax classifier to determine the relation type between entities in the Chinese medical text. Findings–Compared with LSTM and other improved LSTM, the precision of RSig-LSTM rose by 0.96%–5.25%, recall of RSig-LSTM rose by 0.25%–5.25%, F1-Score of RSig-LSTM rose by 0.66%–5.29% and time cost of RSig-LSTM has reduced by 6.97%–33.31%. Originality/value–A local dataset is used to reflect people's physical condition and enhance the practicability of our research. Considering the importance of medical-related entities in medical text research, we use a new formula to calculate the weight of the word in Chinese medical text.

Key Words

LSTM, activation function, relation extraction, Chinese medical text, keyword extraction

* College of Economics & Management, Three Gorges University, Yichang 443002, Hubei, China; e-mail: Lisongpu@foxmail.com
** Hubei Key Laboratory of Intelligent Vision Based Monitoring for Hydroelectric Engineering, Three Gorges University, Yichang 443002, Hubei, China; e-mail: yuxiaosheng@ctgu.edu.cn
*** College of Computer and Information Technology, Three Gorges University, Yichang 443002, Hubei, China; e-mail: 1418083886@qq.com
**** Three Gorges Navigation Authority, Yichang 443002, Hubei, China; e-mail: 16770407@qq.com
Corresponding author: Xiaosheng Yu

Recommended by Xinyu Wu
(DOI: 10.2316/J.2023.206-0850)

1. Introduction

Medical data includes unstructured, semi-structured, and structured data, and medical text is the most important content of medical unstructured data.It is the general term for text data generated by various medical subjects in medical activities, including medical literature, patient electronic medical records, test reports, *etc.* The main problem of relation extraction is to automatically extract the semantic relationship between entities from unstructured text statements and form structured information in a unified format. Specifically, the relation extraction model needs to rely on its own feature extraction ability to predict the possible semantic relationship between entities [1]. Relation extraction has a strong advantage in processing massive data, which is machine translation intelligent search and other tasks improve efficiency and provide a basis for the construction of question answering system and knowledge graph. In medical domains, RE has been applied to drug-gene relationships, disease-symptom relationships, relation extraction for biological pathway construction, *etc.* [2]. Extracting medical knowledge from unstructured medical texts makes it possible to apply knowledge, such as drug discovery, disease prediction, auxiliary diagnosis, and precision medicine. To adapt to the development of emerging information technology, China is stepping up to promote the sharing and application of healthcare data (General Office of the State Council, 2016) [3].

Compared with English, Chinese has unique grammar and sentence structure, but the research on relationship extraction of Chinese medical text starts late, Chinese medical text contains many medical terms, its context is tightly connected, and its content has many forms of expression. So it is difficult to be extracted and study valuable information from the Chinese medical text (Hu *et al.,* 2020) [4]. In Chinese medical texts, various feature generation techniques are used to extract relations from various data, ranging from contextualised word embeddings, *etc.* As the smallest unit to express the subject meaning of a Chinese text, keyword plays a vital role in text classification (Zhao *et al.,* 2017) [5]. Combined with word frequency and feature location, Cao *et al.* (2021) and Miao *et al.* (2020) proposed improved Chinese domain keyword extraction methods to obtain the domain information more reasonably [6], [7]. The next or final step is to select a learning algorithm, such as CNN, RNN, LSTM, *etc.* deep learning algorithm or even language model-based methods, on the features to identify the relations. The deep learning algorithm can alleviate the drawbacks of the relationship extraction model based on classical features to a certain extent, that is, the selection of features is completed by model training, and there is no artificial subjective factor intervention, so the cumulative error is less, which makes the deep learning method more accurate than the classical relationship extraction method based on feature engineering in many occasions [8].

The Tanh activation function of the LSTM model widely used in Chinese medical relationship extraction faces a vanishing gradient problem, and its convergence rate is very slow. This paper proposes a method based on RSigELUS-LSTM (RSig-LSTM) for extracting the relation between entities in Chinese medical text. RSigELUS activation function in RSig-LSTM overcomes the vanishing gradient problem and the negative region problem [9]. This paper uses the bidirectional encoder representation from transformers (BERT) model to obtain vectorisation of health medical texts, and bidirectional RSig LSTM combined with attention mechanism is used to optimise vector features. The Softmax classifier is used to output the probability of relation category between entities in Chinese medical text. In general, the main contributions are as follows.

1. In preprocessing, we use a new formula for calculating weight to extract keywords in Chinese medical text.
2. We use the RSigELUS activation function to improve LSTM. The improved LSTM can overcome the vanishing gradient problem and speed up the convergence rate.
3. The method proposed in this paper performs well in relation extraction tasks of two Chinese medical texts Yidu-S4k and local datasets.

The remainder text of this paper is organised under the following sections. The related works reported on text processing methods are discussed in detail in Section 2. The method of relation extraction used to handle the work is presented in Section 3. In Section 4, experimental datasets and experimental results are discussed in detail. The discussion is presented in Section 5, and the conclusion is presented in Section 6.

2. Related Work

As early as 2010, machine learning was used to process text. The maximum entropy model combined diverse lexical, syntactic, and semantic features to process text and got competitive results. Few studies considered close contextual characteristics of the text. After 2010, the neural network models were gradually applied to text processing. RNN could thoroughly combine contextual text to process information (Mikolov *et al.*, 2013) [10], but it was hard for RNN to ultimately preserve the long-distance features. The variants of RNN (GRU, LSTM) could alleviate the long-distance dependence problem of RNN according to the gating mechanism, so they have been widely recognised in text processing. Guan *et al.* (2019) used bidirectional long–short-term memory (Bi-LSTM) model to extract textual features. BiLSTM could capture bidirectional semantics and obtain textual information more accurately [11]. To reduce the adverse effects of noise, researchers began using LSTM combined with other deep-learning methods to process textual information. Peng *et al.* (2016) and Zhu *et al.* (2018) proposed the Attention-LSTM (ATT-LSTM) model, which used an attention mechanism to assign reasonable weights for features and could effectively highlight essential information [12], [13].

Greff *et al.* (2017) compared the performance of eight improved LSTM models and found that activation functions were one of the most critical parts of the gating structure of LSTM. The output of the Tanh activation function in LSTM is between -1 and 1, but it has the risk of a vanishing gradient. ReLU activation function can overcome the vanishing gradient problem and is easier to be calculated than other activation functions. Le *et al.* (2015) used the ReLU activation function instead of the Tanh activation function in RNN and got as good results as LSTM. However, the output of the ReLU activation function is zero in the negative region and may appear as the phenomenon of neuron death [14]. Clevert *et al.* (2015) proposed the ELU activation function based on the ReLU activation function [15]. ELU activation function uses an exponential function in the negative range. It can reduce output offset and smooth gradient change. Qian *et al.* (2016) proposed a Softplus activation function equivalent to a smooth version of the ReLU activation function [16].

Moreover, it was well-matched to the response function of leaky integrate-and-fire neurons and combined convolutional neural network (CNN) to get great experimental accuracy. Ramachandran *et al.* (2017) proposed the Swish activation function, which combined the advantages of Sigmoid and ReLU activation functions and worked out better than the ReLU activation function on deep-network models [17]. Ye *et al.* (2016) improved the activation function of LSTM and GRU. The improved models reduced the adverse effects of vanishing gradient and improved learning performance [18]. Xu *et al.* (2019) proposed the ArcReLU activation function, an improved function composed of the ReLU activation and Arctan activation functions. It could significantly speed up the back-propagation neural network's training rate and reduce the training error [19]. Xu *et al.* (2021) used the ArcReLU activation function to replace the Tanh activation function of LSTM, which effectively improved accuracy and convergence rate [20]. Kiliarslan *et al.* (2021) proposed the RSigELU activation function, an improved piecewise function composed of double-parameter, Sigmoid, ReLU, and ELU activation functions [9]. It thoroughly used the advantages of piecewise functions and could be effective in the positive, negative, and linear activation regions. The performance of RSigELU was better than other activation functions in several datasets.

In particular domain text research tasks, domain entities are essential for processing text. Frunza *et al.* (2010) used the Naive Bayesian model (NBM) and support vector machine (SVM) to classify the semantic relationship between disease treatment, prevention, and side effects, which can obtain more accurate results than previous studies combining biomedical literature and clinical medical knowledge, but machine learning models had a dependence on manual features [21]. Sahu *et al.* (2016) used CNN to extract traditional Chinese medical relations, which could learn features automatically and reduce the dependency on manual feature engineering [22]. Changfan Zhang *et al.* (2017) studied the effect of different activation functions and the number of hidden layer neurons on recognition performance. The comparison showed that the learning speeds of the ELM and GA-BP were significantly faster, whereas the accuracy rates remained sufficiently high [23]. CNN might inevitably lose features in the pooling process, and RNN could better express time-series features in long text. Zhang *et al.* (2018) used the advantages of CNN and RNN to construct the CNN-RNN model to classify entities in a biomedical text [24]. Guan *et al.* (2019) used Bi-LSTM to extract the relations between entities in the clinical text [11]. Zhang *et al.* (2020) used a bidirectional gate recurrent unit (Bi-GRU) combined with a double attention mechanism to reduce the impact of mislabeling relations between medical entities [25]. Devlin *et al.* (2019) proposed a BERT language model, which could better express polysemy than Word2Vec and make features contain multiple characteristics [26]. Zhang *et al.* (2019) proposed the BERT-BiLSTM-CRF model to identify entities in Chinese electronic medical records. BERT was used to vectorise features in the preprocessing and could alleviate the influence of Chinese polysemy [27]. Du *et al.* (2020) combined BiLSTM with the attention mechanism to determine the category of traditional Chinese medicine text, and the attention mechanism contributed significantly to improving the accuracy of classification [28].

3. Methods

3.1 Preprocessing

3.1.1 Architecture of Preprocessing

The preprocessing of Chinese medical text is shown in Fig. 1. First, we remove a large number of useless symbols in Chinese medical text and use the jieba tool to separate text into words. Then, we use the m-tf-idf algorithm to combine word-length-weight, word-span-weight, and word-class-weight to calculate the final weight of a word in Chinese medical text. The extracted keywords based on final weight are used as entities for extracting relations. Medical experts mark relations between entities according to defined categories of relation. Finally, processed data is divided into train data and test data in the proportion of 8:2.

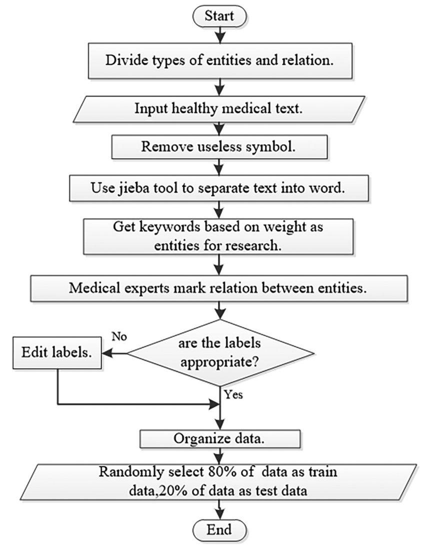

Figure 1. Preprocessing of Chinese medical text.

3.1.2 Keywords Extraction

In text processing, the jieba tool is frequently used to extract keywords. The term frequency-inverse document frequency (tf-idf) algorithm in the jieba tool is used to assign a corresponding word weight according to the word's importance in a document. The jieba tool gets keywords according to calculated weight. The formula of the tf-idf algorithm is shown in (1).

$$w_{a,b} = tf_{a,b} * idf_{\mathrm{a}} = tf_{a,b} * \log\left(\frac{N_{\mathrm{all}}}{df_{a} + 1}\right) \tag{1}$$

$tf_{a,b}$ indicates the frequency of word a in document b, df_{a} indicates the number of documents containing the word a, and N_{tfall} indicates the number of comprehensive documents. The jieba tool selects words with a large output of tf-idf as keywords. Usually, common word (such as "可以" (can), "需要" (need), *et al.*) appears in each document with a specific frequency, the $tfdf_{a}$ of the common word is large, and the tf-idf of the common word is little. If the word appears multiple times in a document and appears fewer times in comprehensive documents, the $tf_{a,b}$ of the word is large, the df_{a} of the word is tiny, and the tf-idf of the word is large. The tf-idf algorithm reduces interference of high-frequency common words for extracting keywords, but it is difficult to highlight the importance of a medical-related word in medical text. This paper proposes the m-tf-idf algorithm combined with word-length-weight, word-span-weight, and word-class-weight to distribute the appropriate weight. It can improve performance to recognise a medical-related word and reduce the adverse effects of a meaningless word in Chinese medical text.

This paper proposes a medical-tf-idf (m-tf-idf) algorithm to measure the importance of a word in Chinese medical text. The formula of the m-tf-idf algorithm is as follows:

$$\begin{aligned}\mathrm{m\_tf\_idf} &= (tf\,(\mathrm{word}|\mathrm{all}) + tf\,(\mathrm{word}|\mathrm{medical})) * \mathrm{idf}_{\mathrm{word}} \\ &= (tf\,(\mathrm{word}|\mathrm{all}) + tf\,(\mathrm{word}|\mathrm{medical})) \\ &\quad * \log\left(\frac{\mathrm{N}_{\mathrm{all}}}{df_{\mathrm{word}} + 1}\right)\end{aligned} \tag{2}$$

$tf\,(\mathrm{word}|\mathrm{all})$ indicates the frequency of the word appearing in the current document, $tf(word|medical)$

indicates the frequency of the word appearing in all medical-related words of the current document, N_{all} indicates the number of comprehensive documents, the df_{word} indicates the number of documents contains the word. The outputs of the tf-idf algorithm in different ranges are added as the m-tf-idf output of the word. The output of the m-tf-idf algorithm is used as one of the evaluation criteria for measuring the importance of the word in Chinese medical text. Frequent common word (such as "需要" (need), "可以" (can), *et al.*) appears in each document with a specific frequency, the tf (word|all) of the frequent common words is large, the df_{word} of the frequent common words approaches to N_{all}, m-tf-idf of the frequent common word is little. The medical-related word appears fewer times in comprehensive documents. When calculating the output of tf, adding tf (word|medical) can increase the frequency of the medical-related words (such as "高血压" (hypertension), "冠心病"(coronary heart disease), *et al.*) and achieve the purpose of highlighting a medical-related word. When calculating the m-tf-idf of the medical-related words, the df_{word} is tiny; the m-tf-idf is large. There are some common words (such as "睡眠情况" (sleep status), "饮食" (eating and drinking), *et al.*) related to medical research; its df_{word} is greater than df_{word} of a medical-related word, its output of m-tf-idf is between frequent common word and medical-related word.

Compared with familiar words, the number of medical-related words ranges from two to seven. Because the longer the length of the word, the more information it contains. The formula of word-length-weight is shown in (3).

$$\text{word\_length\_weight} = \frac{\text{word\_length}}{\text{word\_length} + 5} \tag{3}$$

Word is calculated by using the tf-idf algorithm if it frequently appears in text, which will become a keyword. The more sentences the word contains, the more textual information the word contains. We propose a formula of the word-span-weight (4) to ensure the globality of keywords. $\text{Sentence\_number}_{\text{word}}$ indicates the number of sentences containing the word, and $\text{sentence\_number}_{\text{all}}$ indicates the number of sentences in the current text.

$$\text{word\_span\_weight} = \frac{\text{sentence\_number}_{\text{word}}}{\text{sentence\_number}_{\text{all}}} \tag{4}$$

The part of speech in Chinese text can be divided into nouns, verbs, quantifiers, adverbs, *et al.* Nominal phrases, verbal phrases, and adjectival phrases in the Chinese medical text contain more information than other phrases. Therefore, nominal phrases are the most important, followed by verbal phrases and adjectival phrases, *et al.* The weight of words of different parts of speech (Table 1) is used to distinguish their importance for research.

According to the above formulae, the formula of final weight is shown in (5).

$$\begin{aligned}\text{all\_weight} = {} & (\text{m\_tf\_idf} + \text{word\_length\_weight}) \\ & *\text{word\_span\_weight} \\ & *\text{word\_class\_weight}\end{aligned} \tag{5}$$

Table 1
The Weight of Part of Speech Word

Part of Speech Word	**Weight**
Noun	0.3
Verb	0.2
Adjectives	0.15
Adverbs	0.1
Numerals	0.1
Quantifiers	0.1
Other	0.05

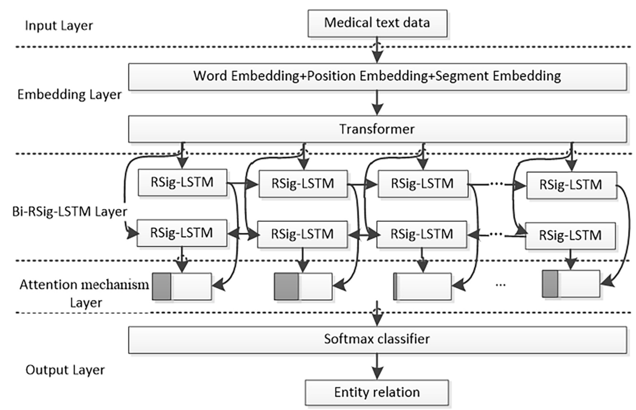

Figure 2. The architecture of RSig-LSTM.

3.2 Architecture of RSig-LSTM

This paper proposes a method based on RSig-LSTM to extract the relation between entities in Chinese medical text. RSig-LSTM model consists of five parts: input layer, embedding layer, Bi-RSig-LSTM layer, attention mechanism layer, and output layer. The architecture of RSig-LSTM is shown in Fig. 2.

(The black component in the box of the attention mechanism layer indicates the importance of the current word)

3.3 Embedding Layer

The Google team proposed the BERT language model based on existing language models, such as Word2Vec and generative pre-training (GPT), composed of multi-layer transformer language architecture. With enormous corpus and supercomputing power, the BERT model achieved excellent performance in textual processing tasks (Devlin *et al.,* 2019) [26]. Embedding vectors of the BERT model are composed of word embeddings, position embeddings, and segment embeddings. BERT enriched information of vector representation and was conducive to subsequent tasks. The context of the Chinese medical text is closely related, and Chinese words can express different meanings in different sentences. We use BERTBase to get the embedding vector (network layer = 12, hidden

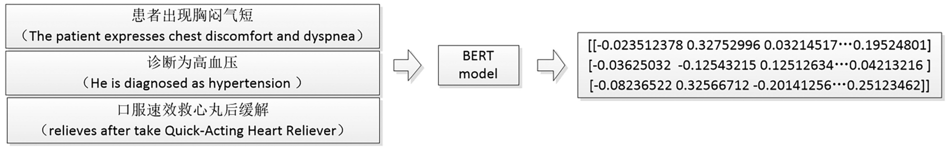

Figure 3. BERT processing of example sentence.

layer dimension = 768, multi-head attention = 12, overall parameters = 110 M). Take sentence ”患者出现胸闷和气短,诊断为高血压,口服速效救心丸后缓解” (The patient expresses chest discomfort and dyspnoea. He is diagnosed as hypertension and relieves after taking Quick-Acting Heart Reliever) as an example. According to the input, the BERT model is used to get vectors. BERT processing of example sentences is shown in Fig. 3.

3.4 Bi-RSig-LSTM Layer

3.4.1 RSigELUS Activation Function

RSigELUS activation function (6) can operate in three active regions (positive, linear, and negative) and can overcome vanishing gradient and negative region problems (Kiliarslan *et al.*, 2021)

$$f(x) = \begin{cases} x\left(\frac{1}{1+e^{-x}}\right)\alpha + x & x > 1 \\ x & 0 \le x \le 1 \\ \beta\left(e^{x} - 1\right) & x < 0 \end{cases} \tag{6}$$

RSigELUS activation function requires double parameters. The parameters of α and β refer to the slope coefficients; α controls the position region, and β controls the negative region. According to the experimental results of Kiliçarslan et al. (2021), the experimental effect is best when $\alpha = 0.5$ and $\beta = 0.2$. So we use $\alpha = 0.5$ and $\beta = 0.2$ as our initial parameters for fine-tuning. RSigELUS function is continuous and derivable at the origin. From the derivative of the RSigelus activation function in (7), $f'(x)$ is always greater than zero. According to the definition of the derivative, it proves RSigELUS activation function is a monotone-increasing function.

$$f'(x) = \begin{cases} \frac{-\alpha x}{(e^{x}+1)^2} + \frac{\alpha x - \alpha}{e^{x}+1} + \alpha + 1 & x > 1 \\ 1 & 0 \le x \le 1 \\ \beta e^{x} & x < 0 \end{cases} \tag{7}$$

RSigELUS activation function has three characteristics. 1) RSigELUS activation function overcomes negative region problem and uses $\beta\left(e^{x}-1\right)$ to calculate in the negative region. 2) It is not saturated in the positive region, alleviating the vanishing gradient problem of the Tanh activation function. 3) Because the RSigELUS activation function is monotone-increasing, it has a faster convergence rate than other traditional activation functions. Figure 4 shows behaviours of traditional activation functions and RSigELUS activation functions.

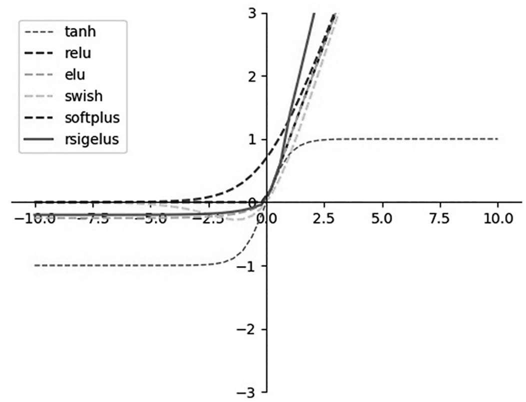

Figure 4. Behaviours of some activation functions.

3.4.2 RSig-LSTM

LSTM is a variant of RNN, which can alleviate long-distance dependence and vanishing gradient problems. LSTM comprises a forget gate, an input, and an output gate. It uses Sigmoid and Tanh activation functions to control the information stock of each gating unit (Hochreiter *et al.*, 1997) [29]. The Sigmoid activation function is in the range from 0 to 1. All memories are forgotten when the output of the Sigmoid activation function is zero; all memories are saved when the output of the Sigmoid activation function is one. The Sigmoid activation function is the most suitable compared with other activation functions. Tanh activation function is mainly used to control information in the control gating unit. However, it faces a vanishing gradient problem, leading to low training efficiency. Therefore, this paper uses the RSigELUS activation function instead of the Tanh activation function to control the information output and improve the convergence rate and accuracy of classification. The internal structure of RSig-LSTM is shown in Fig. 5.

Equation (8)–(12) are calculation formulae of RSig-LSTM. To obtain the hidden vector h_t at t moment, we need to calculate cell state c_t.

$$c_t = c_{t-1} \cdot f_t + i_t \cdot RSigelus\left(W_a h_{t-1} + U_a x_t + b_a\right) \tag{8}$$

x_t is input at t moment, c_{t-1} is cell state at t-1 moment, h_{t-1} is the hidden vector at t-1 moment, f_t represents forget gate and is used to calculate the forgetting probability of cell state at t-1 moment, i_t represents output gate and combines h_{t-1} and x_t to form input at t moment.

$$f_t = \sigma\left(W_f h_{t-1} + U_f x_t + b_f\right) \tag{9}$$

$$i_t = \sigma\left(W_i h_{t-1} + U_i x_t + b_i\right) \tag{10}$$

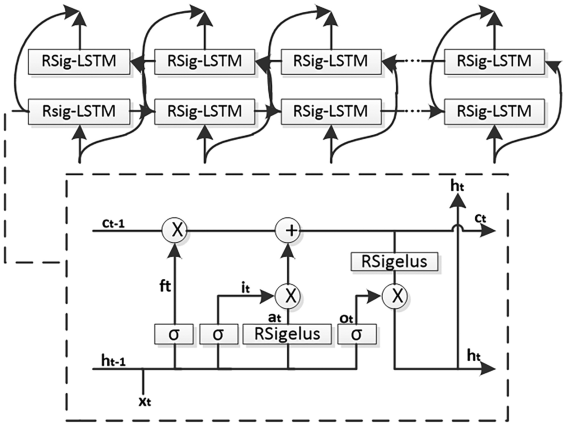

Figure 5. The internal structure of RSig-LSTM.

RSigELUS activation function processes c_t combined with o_t to obtain hidden vector h_t at t moment.

$$o_t = \sigma\left(W_o h_{t-1} + U_o x_t + b_o\right) \tag{11}$$

$$h_t = o_t \cdot RSigelus(c_t) \tag{12}$$

Because unidirectional LSTM can only sequentially process text. To obtain past and future information, combine forward and reverse RSig-LSTM to obtain the final output vector h_t (13) in the Bi-RSig-LSTM layer.

$$h_t = \overrightarrow{h_t} \oplus \overleftarrow{h_t} = RSig\_LSTM(x_t, \overrightarrow{h_{t-1}}) \oplus RSig\_LSTM(x_t, \overleftarrow{h_{t-1}}) \tag{13}$$

Take sentence “患者出现胸闷和气短,诊断为高血压,口服速效救心丸后缓解” (The patient expresses chest discomfort and dyspnoea. He is diagnosed as hypertension and relieves after taking Quick-Acting Heart Reliever) as an example. In Bi-RSig-LSTM layer, “患者(patient)”, “出现(express)”, “胸闷(chest discomfort)”, “气短(dyspnoea)”, “诊断(diagnose)”, “高血压(hypertension)”, “口服(take)”, “速效救心丸(Quick-Acting Heart Reliever)”, “缓解(relieve)” express as {hi1,hi2,hi3,hi4,hi5,hi6,hi7,hi8,hi9} and {hj1,hj2,hj3,hj4,hj5,hj6,hj7,hj8,hj9}. Them are spliced into {h1,h2,h3, h4,h5,h6,h7,h8,h9}. Bi-RSig-LSTM processing of example sentences is shown in Fig. 6.

3.5 Attention Mechanism Layer

In the Chinese medical text, the contribution of each word is different. If unified weight is assigned to a word, some critical information will be omitted, and experimental accuracy will be reduced. The attention mechanism model can adjust the weight of entities according to the degree of contribution to research; it can quickly get and focus on processing valuable information. Take sentence “患者出现胸闷和气短,诊断为高血压,口服速效救心丸后缓解” (The patient expresses chest discomfort and dyspnoea. He is diagnosed as hypertension and relieves after taking Quick-Acting Heart Reliever) as an example. We use the attention mechanism to assign corresponding weight {a1,a2,a3, a4,a5,a6,a7,a8,a9} to vectors, combining weight a_t and vector h_t to new vector s_t. The Softmax function is used to get predicted relation categories between entities in Chinese medical text. The Attention mechanism processing of an example sentence is shown in Fig. 7.

4. Resluts

4.1 Datasets

After searching a large number of medical electronic healthy records, we divided categories of medical-related entities and categories of relation in Chinese medical text. After consulting medical experts, we adjust standards of classification standards. Medical-related entities are divided into 12 categories, and the relation of medical-related entities is divided into 10 categories. The details of entity type are shown in Table 2. The details of the relation type are shown in Table 3.

Each relation in datasets consists of a specific type of entity pair. The entity pairs with the same type are divided into the same group. There are multiple entities in each sentence. We combine different types of entities and get different entity pairs. The different entity pairs have different relation types. We obtain three entity pairs in each instance for our research. The relation structure between entities of Chinese medical text is shown in Fig. 8.

This paper uses two medical datasets for research. 1) Yidu-S4K dataset: It is manually edited by the ”Yidu-Cloud” medical data intelligent platform based

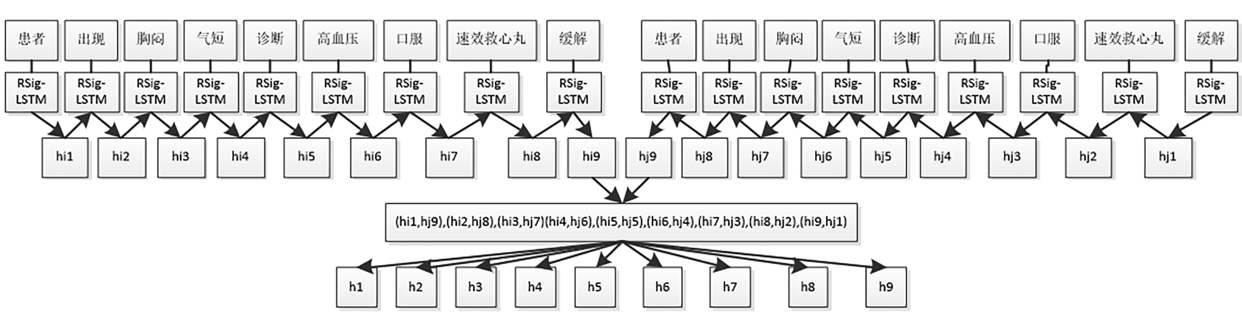

Figure 6. Bi-RSig-LSTM processing of example sentence.

Table 2
The Details of the Entity Type

Entity Type	Examples of Entity
疾病(disease)	冠心病(coronary heart disease)、高血压(hypertension)
症状(symptom)	胸闷(chest discomfort)、气短(dyspnoea)
药品(medicine)	阿莫西林(amoxicillin)、复方丹参片(compound Salvia tablets)
检查方式(exam)	测量血压(blood pressure measurement)、抽血化验(blood test)
治疗方式(cure)	心脏搭桥手术(heart bypass surgery)、透析(dialysis)
单位定量(unit)	毫克(mg)、毫升(ml)
身体部位(body part)	肾脏(kidney)、左眼(left eye)
科室(department)	妇产科(obstetrics and gynecology department)、心血管内科(cardiovascular Internal medicine practice)
观察指标(index)	睡眠状况(sleep status)、排便情况(defecation)
检查治疗状态(condition)	剧痛(sharp pain)、昏迷(coma)
治疗结果(result)	康复(recovery)、恶化(deteriorate)
病变(lesions)	积液(effusion)、癌变(canceration)

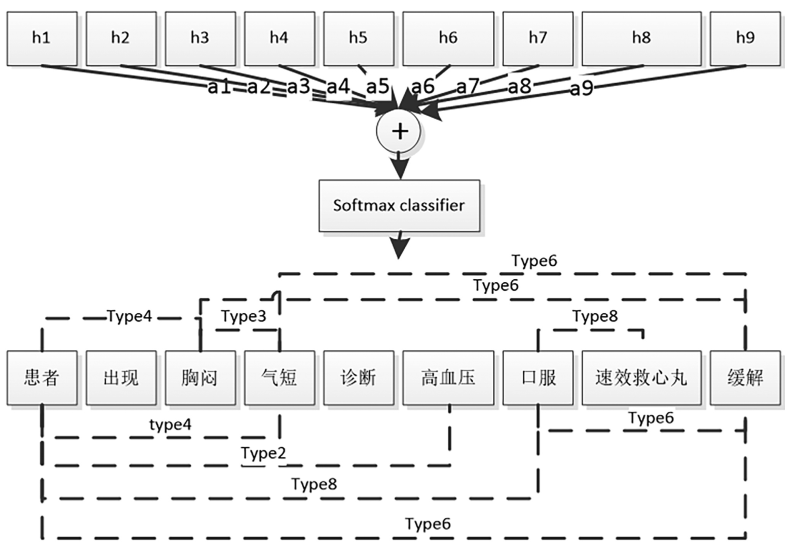

Figure 7. Attention mechanism processing of example sentence.

on the distribution of actual medical records, with a total of 8,000 records, including patient symptoms, previous medical history, examination, and treatment, *et al.* Many researchers use this dataset because it is the network platform dataset. We can obtain effective experimental results. 2) Local dataset: It is desensitisation data of patients with hypertension in a city's Centres for Disease Control from 2017 to 2018, with a total of 5,830 records, including previous medical history, patient symptoms, patient recovery situation, *et al.* Using local data can genuinely reflect people's reality and make experimental results more convincing. The proportion of different categories in datasets is shown in Fig. 9.

4.2 Experimental Results

This paper uses macro-average to reflect the performance of the model. The precision, recall, and F1-Score are used as experimental evaluation criteria. According to precision, recall, and F1-Score of each iteration, adjusting experimental parameters and making the model reach the best condition. Experimental parameters in the best condition of the RSig-LSTM model are shown in Table 4. If the loss rate does not decrease compared with the previous epoch of training in the iterative process, it needs to continue five times. If the loss rate also does not change, iteration is terminated. The experimental results of each category are shown in Table 5.

Table 3
The Details of the Relation Type

	Relation Type
Type 1	Unknown(The relationship cannot be defined)
Type 2	诊断情况(diagnostic condition)
Type 3	相同类别(same type)
Type 4	患者症状(clinical manifestation)
Type 5	用量(Specific dosage of examination,treatment and index)
Type 6	治疗结果与表现(Results of treatment)
Type 7	检查与治疗状态(examination and treatment status)
Type 8	检查治疗方式(Examination and treatment of patients)
Type 9	炎症位置与状态(part and state of lesions)
Type 10	观察指标变化情况(Change of indexes)

Table 4
The Experimental Parameters of RSig-LSTM

Parameter	**Value**
BERT vector dimension	768
LSTM vector dimension	128
Attention vector dimension	32
Batch size	16
Dropout-rate	0.5
Optimiser	Adam
Learning rate	0.001

4.3 Comparison of Experimental Results and Analysis

To explore the impact of different modules on the relationship extraction performance of this model, we also designed a group of ablation experiments, controlling a single variable each time to determine its impact on the model. Our model has two primary innovations: 1) new formula for calculating weight to extract keywords (m-tf-idf) and 2) RSigELUS activation function to improve LSTM. So the model is mainly divided into three parts. In addition, it includes an attention mechanism.

As shown in Table 6, when we only use the m-tf-idf weight calculation formula based on LSTM, the model improves precision, recall, and F1 scores, and F1 scores increase by 2.3%. When we remove the attention mechanism, the model F1 scores only decrease by 1%. When we use the m-tf-idf weight calculation formula and the RSigeLUS activation function, the model improves in precision, recall, and F1 scores, and F1 scores increase by 4.4%; this shows that the improvement of the model in relation extraction is mainly due to the module of m-tf-idf and RSigeLUS activation function.

In addition, we compared our model with other models. The GlobalPointer model is commonly used to extract the nearest relationships. J Su (2022) proposed using GlobalPointer to extract entity relationships [30], and Liang J *et al.* (2022) proposed a joint extraction model based on soft pruning and GlobalPointer [31]. We conducted a comparative experiment with the GlobalPointer model. Due to data formatting problems, the experiment used the Chinese medical text entity relationship dataset CHIP-2020-2. The experimental results in Table 7 show that our model achieves better performance.

In this paper, the proposed RSig-LSTM is compared with traditional LSTM (Du *et al.* 2020), LSTM improved by using the ReLU activation function (ReLU-LSTM, Doni *et al.* 2020) [32], LSTM improved by using the ELU activation function (ELU-LSTM, Ji et al. 2019) [33], LSTM improved by using the Swish activation function (Swish-LSTM, Eger at al. 2019) [34] and LSTM improved by using the Softplus activation function (Softplus-LSTM, Ye et al. 2016) [18]. Table 8 shows the precision, recall, F1 score, and time cost of the LSTM module of each model in the experimental dataset. The time cost of the LSTM module of each model can indicate the impacts of different activation functions.

The number of epochs executed by the model is one of the essential criteria to explain its efficiency. Figure 10 shows the iterative process of six models in the Yidu-S4K dataset and the local dataset. In the Yidu-S4K dataset, the ELU-LSTM, and Softplus-LSTM converge at the 10th epoch, followed by Swish-LSTM (11th epoch), ReLU-LSTM (12th epoch), RSig-LSTM (16th epoch), and LSTM (17th epoch) In the local dataset, the ReLU-LSTM converges at the 8th epoch, followed by ELU-LSTM (9th epoch), RSig-LSTM (12th epoch), Softplus-LSTM (16th epoch), Swish-LSTM (17th epoch), and LSTM (20th epoch). In addition to the number of

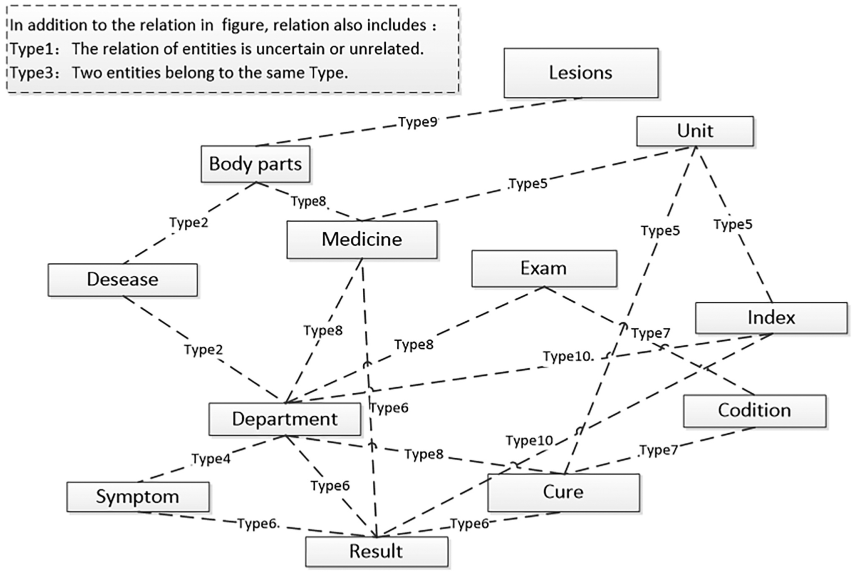

Figure 8. Relation structure between entities of Chinese medical text.

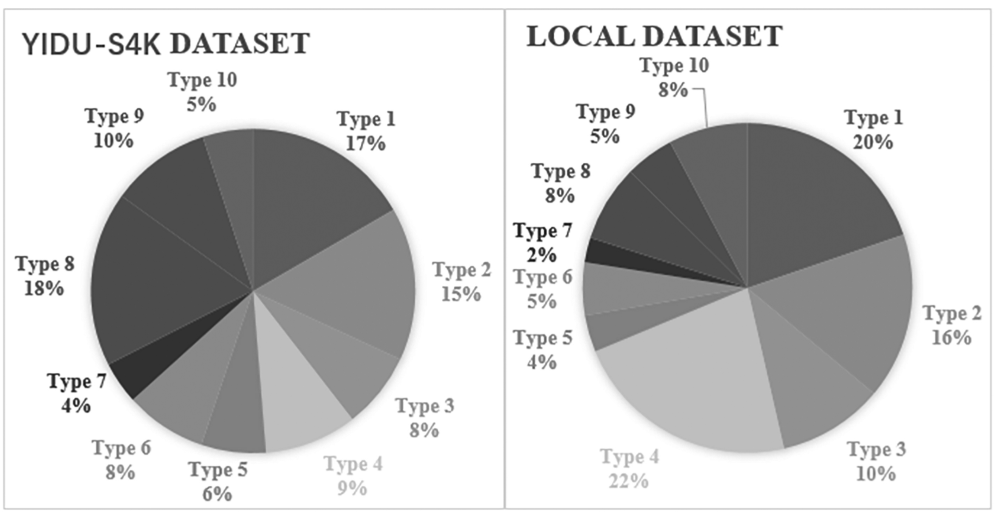

Figure 9. The proportion of different categories in datasets.

epochs, the time cost is necessary to evaluate the model's efficiency. Figure 11 is used to show the time cost of six models.

Compared with LSTM, ReLU-LSTM, and ELU-LSTM, the precision of RSig-LSTM rose by 1.77%–5.25%, recall of RSig-LSTM rose by 0.71%–5.25%, F1-Score of RSig-LSTM rose by 1.37%–5.29% and overall time cost of RSig-LSTM has reduced by 6.97%–33.31%. The evaluation indexes of RSig-LSTM are significantly improved, and the overall time cost of RSig-LSTM is greatly reduced. Compared with Swish-LSTM and Softplus-LSTM, the precision of RSig-LSTM rose by 0.96%–1.86%, recall of RSig-LSTM rose by 0.25%–1.2%, F1-Score of RSig-LSTM rose by 0.66%–1.23%, and overall time cost of RSig-LSTM has reduced by 13.53%–32.87%. The evaluation indexes of RSig-LSTM are slightly improved, but the overall time cost of RSig-LSTM is greatly reduced. Considering evaluation indexes and overall time cost, RSig-LSTM performs better than other models.

Discussion

The current work partly solves the intractable problems in researching health medical texts, such as complex structures, keywords in professional fields, and synonymous expressions. In addition to the meaning of entities, the Chinese medical text also contains the implicit relation information between entities. The relation extraction method is crucial to obtaining more valuable information in Chinese medical texts. In our study, the relation extraction method uses a new weight calculation method to highlight medical keywords and reduce the influence of meaningless words. It can better get medical information and reduce interference to RSig-LSTM. The RSigELUS activation

Table 5
Experimental Results of RSig-LSTM

Yidu-S4K Dataset					**Local Dataset**				
Relation Type	**Precision**	**Recall**	**F1-Score**	**Support**	**Relation Type**	**Precision**	**Recall**	**F1-Score**	**Support**
Type 1	0.7850	0.7393	0.7615	251	Type 1	0.9269	0.8728	0.8990	206
Type 2	0.8458	0.9065	0.8751	243	Type 2	0.9330	0.9474	0.9401	193
Type 3	0.8529	0.7398	0.7923	126	Type 3	0.9845	0.9143	0.9481	147
Type 4	0.8954	0.9257	0.9103	151	Type 4	0.9196	0.9565	0.9377	255
Type 5	0.9083	0.9245	0.9163	101	Type 5	1.0000	0.9444	0.9714	40
Type 6	0.8468	0.8750	0.8607	117	Type 6	0.9480	0.9194	0.9335	61
Type 7	1.0000	0.9710	0.9853	80	Type 7	0.9167	0.9615	0.9386	48
Type 8	0.9031	0.9273	0.9150	272	Type 8	0.9545	0.9770	0.9656	80
Type 9	0.8750	0.8452	0.8598	165	Type 9	0.9556	0.8600	0.9053	53
Type 10	0.9048	0.9326	0.9185	94	Type 10	0.8317	0.9556	0.8894	83
Macro avg	0.8817	0.8787	0.8795	1600	Macro avg	0.9371	0.9309	0.9329	1166
Weighted avg	0.8694	0.8697	0.8689	1600	Weighted avg	0.9331	0.9297	0.9305	1166

Table 6
Results of Ablation Experiments

Model	**Yidu-S4K Dataset**			
	Inputs Used	**Precision**	**Recall**	**F1**
LSTM (without RSig+m-tf-idf)	tf-idf+Att	0.8377	0.8349	0.8353
LSTM (without attention)	m-tf-idf	0.8443	0.8444	0.8434
LSTM (without RSig)	m-tf-idf+Att	0.8608	0.8600	0.8587
RSig-LSTM	m-tf-idf+RSig+Att	0.8817	0.8787	0.8795

Table 7
Comparative Evaluation of GlobalPointer and RSig-LSTM

Model	**CHIP-2020-2 Dataset**		
	Precision	**Recall**	**F1**
GlobalPointer	0.6371	0.5473	0.5888
RSig-LSTM	0.6544	0.5871	0.6055

function improves RSig-LSTM. RSig-LSTM solves the problem of neuronal necrosis, reduces the risk of gradient disappearance, and calculates the cost. Compared with other LSTM algorithms, RSig-LSTM has the lowest overall time cost and faster computing speed. In addition, it has better accuracy, recall rate, and F1 score, showing better computing ability. Therefore, our relational extraction model has proved helpful in health medical texts.

Conclusion

There are many professional terms and complex structures in Chinese medical texts. The implicit relation information between two entities in the Chinese medical text is often complicated. Deep learning relieved the drawbacks of machine learning relying on manual extraction features. Therefore, we use LSTM combined with BERT and attention mechanism to extract medical relations. To alleviate the vanishing gradient problem of the Tanh activation function in LSTM, we propose RSig-LSTM extract relation in Chinese medical text. We consider the particularity of relational entities in Chinese medical texts; the m-tf-idf algorithm is used as the criterion of importance in Chinese medical text combined with word-length-weight, word-span-weight, and word-class-weight. Experimental results on two Chinese medical clinical datasets are used to show that RSig-LSTM is highly effective than other LSTM models. This paper only classifies the question with a single label and does not classify disease

Table 8
Experimental Results of Six Methods

Methods	Yidu-S4K Dataset				Local Dataset			
	Time Cost of LSTM Module (s)	Precision	Recall	F1	Time Cost of LSTM Module (s)	Precision	Recall	F1
LSTM	0.57	0.8377	0.8349	0.8353	0.67	0.9020	0.9107	0.9049
ReLU-LSTM	0.59	0.8535	0.8545	0.8525	0.60	0.9170	0.9243	0.9203
ELU-LSTM	0.63	0.8617	0.8679	0.8634	0.56	0.9208	0.9209	0.9198
Swish-LSTM	0.58	0.8656	0.8749	0.8688	0.56	0.9268	0.9273	0.9268
Softplus-LSTM	0.76	0.8733	0.8683	0.8692	0.64	0.9220	0.9286	0.9247
RSig-LSTM	0.55	0.8817	0.8787	0.8795	0.53	0.9371	0.9309	0.9329

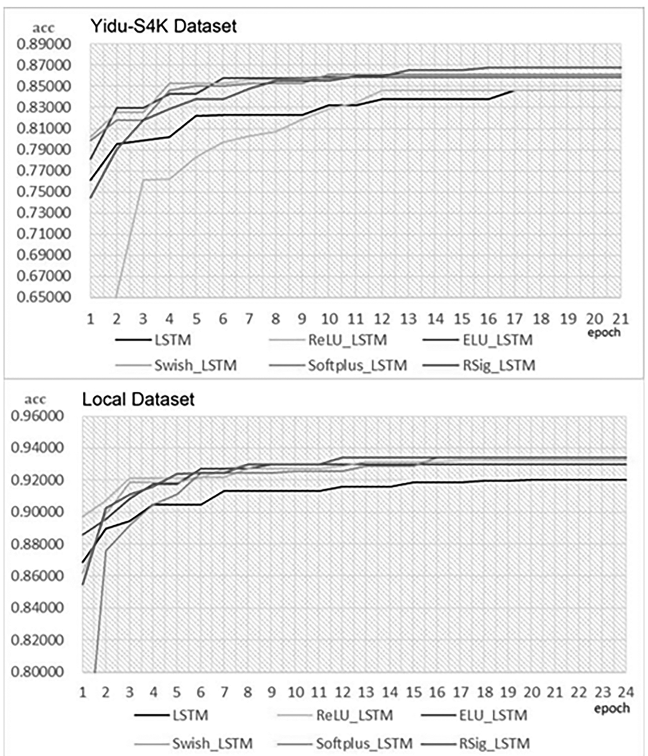

Figure 10. The iterative process of six models in the local dataset.

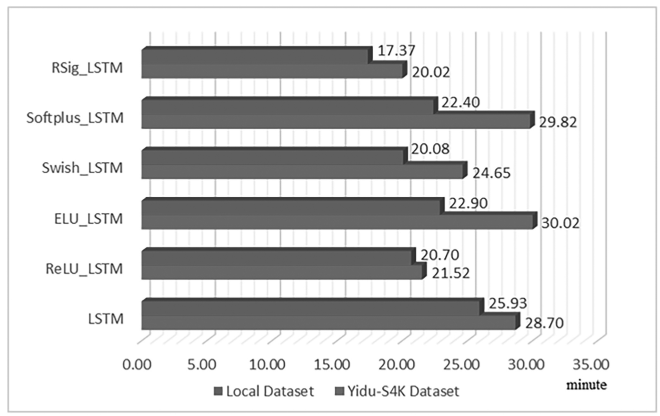

Figure 11. The time cost of six models.

labels in the disease problem. Because of the limitation of experimental conditions, the size of the datasets is small.

References

[1] T. Xie, J.A. Yang, and H. Liu, Chinese entity relation extraction based on multi-feature BERT model, *Computer Systems and Applications*, *30*(5), 2021, 253–261.

[2] B. Priyankar, S. Srinivasan, W.C. Sleeman, J. Palta, R. Kapoor, and P. Ghosh, A survey on recent named entity recognition and relationship extraction techniques on clinical texts, *Applied Sciences*, *11*(8), 2021, 8319.

[3] General Office of the State Council, Guidance of the General Office of the State Council on promoting and standardizing the application and development of health data, https://www.gov.cn/zhengce/content/2016-06/24/content_5085091.htm (accessed Oct. 2022).

[4] J.H. Hu, W.Q. Zhao, and A. Fang, Research on clinical text processing and knowledge discovery method based on medical big data, *China Digital Medicine*, *15*(7), 2020, 13–25.

[5] J.S. Zhao, Q.M. Zhu, and G.D. Zhou, Review of research in automatic keyword extraction, *Journal of Software*, *28*(9), 2017, 2431–2449.

[6] Y.Q. Cao, W.P. Sheng, and H.X. Zhou, Research on news keyword extraction based on TF-IDF-MP algorithm, *Journal of East China Jiaotong University*, *38*(1), 2021, 122–130.

[7] C. Miao, Z. Cao, and Y.C. Tam, Keyword-attentive deep semantic matching, 2020, *arXiv:2003.11516* (accessed Mar. 2021).

[8] X.P. Wu, Q. Zhang, F. Zhao, and L. Jiao, Entity relation extraction method for guidelines of cardiovascular disease based on bidirectional encoder representation from transformers, *Journal of Computer Applications*, *41*(1), 2021, 145.

[9] S. Kiliarslan and M. Celik, RSigELU: A nonlinear activation function for deep neural networks, *Expert Systems With Applications*, *174*, 2021, 114805.

[10] T. Mikolov, I. Sutskever, and K. Chen, Distributed representations of words and phrases and their compositionality, 2013, *arXiv:1310.4546v1* (accessed Oct. 2021).

[11] P.J. Guan and C.P Cao, Clincal text entity relationship extraction based on BiLSTM, *Computer Engineering and Software*, *5*(1), 2019, 159–62.

[12] Z. Peng, S. Wei, and J. Tian, Attention-based bidirectional long short-term memory networks for relation classification, *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics*, Berlin, 2016, 207–212.

[13] X.J. Zhu, H.L. Li, and J.S. Zhou, An improved attention-based LSTM feature selection model, *Journal of Beijing Institute of Machinery*, *33*(122), 2018, 57–62.

[14] Q.V. Le, N. Jaitly, and G.E. Hinton, A simple way to initialize recurrent networks of rectified linear units, 2015, *arXiv:1504.00941* (accessed Apr. 2022).

[15] D.A. Clevert, T. Unterthiner, and S. Hochreiter, Fast and accurate deep network learning by exponential linear units(ELUs), 2015, *arXiv:1511.07289* (accessed Feb. 2022).

[16] L. Qian and S. Furber, Noisy softplus: A biology inspired activation function, in *Proc. International Conf. on Neural Information Processing, Neural Information Processing*, Kyoto, 2016, 405–412.

[17] P. Ramachandran, B. Zoph, and Q.V. Le, Searching for activation functions, 2017, *arXiv:1710.05941* (accessed Oct. 2021).

[18] X.Z. Ye, F.F. Tao, and R.Z. Qi, Inprovement on activation functions of recurrent neural network architectures, *Computer and Modernization*, 2016, 29–33.

[19] B.J. Xu and F.F. Xu, Optimization of activation function in neural network based on ArcReLU function, *Journal of Data Acquisition and Processing*, *34*(3), 2019, 517–529.

[20] F.F. Xu and B.J. Xu, Research on matching resumes and position based on Arc-LSTM, *Journal of Shangdong University (Natural Science)*, *56*(1), 2021, 83–90.

[21] O. Frunza and D. Inkpen, Extraction of disease-treatment semantic relations from biomedical sentences, *Proc. of the 2010 Workshop on Biomedical Natural Language Processing*, Stroudsburg, PA, 2010, 91–98.

[22] S.K. Sahu, A. Anand, and K. Oruganty, Relation extraction from clinical texts using domain invariant convolutional neural network, *Proc. of the 15th Workshop on Biomedical Natural Language Processing*, Berlin, 2016, 206–215.

[23] C. Zhang, X. Cheng, J. He, and G. Liu, Automatic recognition of adhesion states using an extreme learning machine, *International Journal of Robotics and Automation*, *32*(2), 2017, 194–200.

[24] Y. Zhang, H. Lin, and Z. Yang, J. Wang, S. Zhang, Y. Sun, and L. Yang, A hybrid model based on neural networks for biomedical relation extraction, *Journal of Biomedical Informatics*, *81*, 2018, 83–92.

[25] Z.C. Zhang, T. Zhou, R.F. Zhang, and M.Y. Zhang, Medical entity relation recognition combining bidirectional GRU and attention, *Computer Engineering*, *46*(514), 2020, 302–308.

[26] J. Devlin, M.W. Chang, and K. Lee, BERT: Pre-training of deep bidirectional transformers for language understanding, 2019, *arXiv:1810.04805v2* (accessed May 2022).

[27] W. Zhang, S. Jiang, and S. Zhao, A BERT-BiLSTM-CRF model for Chinese electronic medical records named entity recognition, *Proc. 12th International Conf. on Intelligent Computation Technology and Automation (ICICTA)*, Xiangtan, 2019, 166–169.

[28] Y. Hui, L. Du, S. Lin, Y. Qu, and D. Cao, Extraction and classification of TCM medical records based on BERT and Bi-LSTM with attention mechanism, *Proc. IEEE International Conf. on Bioinformatics and Biomedicine (BIBM)*, Seoul, 2020, 1626–1631.

[29] S. Hochreiter and J. Schmidhuber, Long short-term memory, *CNeural Computation*, *9*(8), 1997, 1735–1780.

[30] J. Su, GPlinker: Entity-relation joint extraction based on GlobalPointer, 2022, https://kexue.fm/archives/8373 (accessed May 2022).

[31] J. Liang, Q. He, D. Zhang, and S. Fan, Extraction of joint entity and relationships with soft pruning and GlobalPointer, *Applied Sciences*, *12*(13), 2022, 6361.

[32] A. Doni and T. Sasipraba, Lstm-RNN based approach for prediction of dengue cases in India, *Ingénierie des Systèmes d'Information*, *25*(3), 2020, 327–335.

[33] X.Q. Ji, *Short term electricity price forecasting based on deep learning in electricity market*, (Beijing: North China Electric Power University, 2019).

[34] S. Eger, P. Youssef, and I. Gurevych, Is it time to swish? Comparing deep learning activation functions across NLP tasks, 2019, *arXiv:1901.02671v1* (accessed Jan. 2022).

Biographies

Songpu Li is currently a Doctoral candidate in information management and information system with the School of Economics and Management, Three Gorges University. His research interests are medical text natural language processing and medical knowledge graph application.

Ruxin Gong received the master's degree in computer science from the College of Computer Science and Information Technology, Three Gorges University in 2022. Her research interests are medical big data, natural language processing, and text analysis.

Xiaosheng Yu received the Ph.D. degree in information science from Wuhan University in 2007. Since 2010, he has been an Associate Professor with the College of Computer Science and Information Technology of China, Three Gorges University. His research interests are big data analysis and information fusion.

Jiaheng Li received the master's degree in engineering from the China Three Gorges University in 2011. He is a Senior Engineer with the Three Gorges Navigation Authority. His research interests are intelligent navigation organization and big data analysis.