



# Extraction of entity relations from Chinese medical literature based on multi-scale CRNN

Tingyin Chen<sup>1,2#^</sup>, Xuehong Wu<sup>3,4#^</sup>, Linyi Li<sup>3,4</sup>, Jianhua Li<sup>3</sup>, Song Feng<sup>1,2</sup>

<sup>1</sup>Department of Network and Information, Xiangya Hospital, Central South University, Changsha, China; <sup>2</sup>National Clinical Research Center for Geriatric Disorders (Xiangya Hospital), Changsha, China; <sup>3</sup>Hunan Creator Information Technology Co. Ltd, Changsha, China; <sup>4</sup>School of Computer Science and Engineering, Central South University, Changsha, China

**Contributions:** (I) Conception and design: T Chen, S Feng; (II) Administrative support: J Li, S Feng; (III) Provision of study materials or patients: T Chen, X Wu, L Li; (IV) Collection and assembly of data: T Chen, X Wu, L Li; (V) Data analysis and interpretation: T Chen, X Wu; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

<sup>#</sup>These authors contributed equally to this work.

**Correspondence to:** Song Feng. Network Information Center, Xiangya Hospital, Central South University, Xiangya Road, Changsha 410008, China. Email: fs205@sina.com.

**Background:** Entity relation extraction technology can be used to extract entities and relations from medical literature, and automatically establish professional mapping knowledge domains. The classical text classification model, convolutional neural networks for sentence classification (TEXTCNN), has been shown to have good classification performance, but also has a long-distance dependency problem, which is a common problem of convolutional neural networks (CNNs). Recurrent neural networks (RNN) address the long-distance dependency problem but cannot capture text features at a specific scale in the text.

**Methods:** To solve these problems, this study sought to establish a model with a multi-scale convolutional recurrent neural network for Sentence Classification (TEXTCRNN) to address the deficiencies in the 2 neural network structures. In entity relation extraction, the entity pair is generally composed of a subject and an object, but as the subject in the entity pair of medical literature is always omitted, it is difficult to use this coding method to obtain general entity position information. Thus, we proposed a new coding method to obtain entity position information to re-establish the relationship between subject and object and complete the entity relation extraction.

**Results:** By comparing the benchmark neural network model and 2 typical multi-scale TEXTCRNN models, the TEXTCRNN [bidirectional long- and short-term memory (BiLSTM)] and TEXTCRNN [double-layer stacking gated recurrent unit (GRU)], the results showed that the multi-scale CRNN model had the best F1 value performance, and the TEXTCRNN (double-layer stacking GRU) was more capable of entity relation classification when the same entity word did not belong to the same entity relation.

**Conclusions:** The experimental results of the entity relation extraction from *Pharmacopoeia of the People's Republic of China—Guidelines for Clinical Drug Use—Volume of Chemical Drugs and Biological Products* showed that entity relation extraction could effectively proceed using the new labeling method. Additionally, compared to typical neural network models, including the TEXTCNN, GRU, and BiLSTM, the multi-scale convolutional recurrent neural network structure had advantages across several evaluation indicators.

**Keywords:** Medical literature; entity relation extraction; convolutional recurrent neural network

Submitted Feb 16, 2022. Accepted for publication Apr 07, 2022.

doi: 10.21037/atm-22-1226

**View this article at:** <https://dx.doi.org/10.21037/atm-22-1226>

<sup>^</sup> ORCID: Tingyin Chen, 0000-0002-2365-8281; Xuehong Wu, 0000-0003-4861-5648.

## Introduction

Entity relation extraction is an important task in natural language processing (NLP), and its main purpose is to capture semantic relations existing in known entity pairs in unstructured text. As intelligent diagnosis and inquiry and other artificial intelligence technologies are realized through inferences based on mapping knowledge domains, the significance of entity relation extraction is that it can identify the entity relations between knowledge entities for the mapping knowledge domains. If the entity relations in a medical literature cannot be effectively extracted, the knowledge integrity of the established mapping knowledge domains will be affected, and wrong conclusions may be drawn subsequent inferences. Thus, it is necessary to select a suitable neural network model for entity relation extraction.

There are 2 main types of classical neural network models; that is, the convolutional neural network (CNN), and the recurrent neural network (RNN). Like the traditional n-gram model, the CNN is used to capture text features at a specific scale rather than time-sequence characteristics in the text. Conversely, RNN is used to capture the time-sequence characteristics of the text rather than the text features at a specific scale. If only 1 neural network model is applied, it is inevitable that some features of the text shall be missed. This has become an urgent issue that needs to be addressed. Additionally, the incomplete text structure of medical literature is also a severe problem. The object of entity relation extraction is to extract an entity pair composed of a subject and an object. For professional literature, as each chapter has the same semantic background, the omission of the subject has become a common way to represent text. However, for entity relation extraction, the subject omitted always serves as the subject in an entity pair. If the entity pair is missing, the general entity position coding method for entity relation extraction is no longer applicable.

Entity relation extraction technologies are mainly divided into 2 types; that is, methods based on template rules and methods based on feature vectors. In relation to the methods based on template rules, generally, the language specialists first summarize the language features of the entity relations, and then write corresponding entity relation rules to realize entity relation extraction by rule matching (1-4). The methods based on feature vectors can be divided into traditional machine learning and deep learning. In the former, after the feature vectors are established, traditional

machine learning models, such as the maximum entropy classifier (5,6) or the support vector machine (7), may be used to complete the construction of entity relation extraction, and the performance of the model construction mainly depends on the selection or construction of the feature vectors. In relation to the latter, the selection of feature vectors is simple. The corresponding feature vector (i.e., the word vector) (8) is generated simply by embedding words (or expressions) into the specific vector space, and the model performance is mainly affected by the selection of the neural network.

The CNN can sufficiently capture the language features in the word vector (9-11), but the convolutional feature causes a long-distance dependency problem whereby it is difficult to associate words with a long distance between them. To solve the long-distance dependency problem, Lerner *et al.* introduced the RNN (12), in which the neurons are connected end to end (i.e., the input of a nerve cell is the output of its previous nerve cell, and the output of the nerve cell is the input of its next nerve cell). With this structure design, even 2 words with a long distance between them may be associated correspondingly, but as the operation process within the nerve cells is mainly achieved by multiplication, the gradient explosion and vanishing problem is inevitable. To solve this problem, Alt *et al.* adopted the transformer structure neural network (13), whereby, via the application of the self-attention mechanism and the multi-head attention mechanism, not only are 2 words with a long distance between them associated, but the gradient explosion or loss problem is also avoided.

A medical text represents a specific application scenario, so more specific problems may appear in the entity relation extraction. For example, when the support vector machine is used to extract relations (the training process of the support vector machine aims to find the maximum margin hyperplane), misclassification occurs more frequently for some relation types that have not been subdivided. To solve such a problem, relations are subdivided, and then several support vector machines are used for the classification (14). In relation to entity relation extraction from inquiry dialogues, Hassan *et al.* (15) stored the entities in the form of cache, and then adopted the combined extraction method to lower the complexity of  $O(n^4)$  to  $O(n)$ .

Based on the characteristics of the CNN and RNN, many studies in NLP have focused on the combination of the 2 networks. For example, the network models have been connected in the order of CNN+RNN in text classification, features have been captured in the form of

RNN+RNN+CNN in questions and answers (16), and deeper serial information has effectively been obtained by stacking 2 layers of RNN. In terms of entity relation extraction in the medical field, in the RNN+CNN structure, the adding of a pooling layer between 2 neural networks led to the unnecessary features in the results that RNN transmits to CNN being omitted, and only significant features expressing the entity relation semantics being transmitted (17). In relation to the usually neglected syntactic structure, in a previous study, after the shortest dependency path for the syntax in the text was extracted, the CNN+RNN structure was adopted to extract the syntactic features (18).

The methods mentioned above represent simple applications of the CNN; however, the most important application of the convolution kernel in the CNN has not been used. Convolution kernels at different scales capture text features at different scales. In the TEXTCNN (19), multi-scale convolution kernels are used simultaneously to capture features at different scales, and thus its performance is superior to that of simple models with a single convolution kernel. As multi-scale convolution kernels are used, it is inevitable that the model parameters become bigger. A study by Sahu *et al.* showed that with the use of size 4 and size 6 convolution kernels, the performance was similar (20).

Thus, we proposed the multi-scale convolutional recurrent neural network (CRNN) of TEXTCRNN. By reference to the CNN+RNN structure, the TEXTCNN was modified to the CNN+RNN structure in this model. Additionally, 3 scales of convolution kernels were reduced to 2 scales, and thus the model parameters were decreased. By experimenting with the benchmark model and 2 typical multi-scale CRNN models [i.e., the TEXTCRNN (BILSTM) and TEXTCRNN [double-layer stacking gated recurrent unit (GRU)]], we found that the multi-scale CRNN model could be used to capture the time-sequence characteristics of features at different scales in the text.

## Methods

### *Entity relation extraction in the case of a single entity*

In this study, a professional literature in the field of medicine, *Pharmacopoeia of the People's Republic of China—Guidelines for Clinical Drug Use—Volume of Chemical Drugs and Biological Products (Pharmacopoeia)*, served as the experimental data set. The common problem of professional

literature (i.e., that the subject is always omitted) also occurred in this data set. In the text of professional literature, the subject term in a chapter is always regarded as the subject by default, and it is only not omitted when the subject is not the subject term. However, in general entity relation extraction, the entity relations can only be extracted when 2 entities exist in the text at the same time and form an entity pair. In this data set, as only 1 entity existed, the entity pair could not be formed effectively. For this reason, the general method could not be directly adopted to represent the entity position information, and thus the entity relations could not be extracted. Thus, a new entity position representation form was proposed.

If the general method had been used, the representation form would have been as follows:

“XX<sub>(0, -9)</sub> is used for<sub>(1, -8)</sub> eye<sub>(2, -7)</sub> infections<sub>(3, -6)</sub> caused by<sub>(4, -5)</sub> sensitive<sub>(5, -4)</sub> fungi<sub>(6, -3)</sub>, <sub>(7, -2)</sub> such as<sub>(8, -1)</sub> fungal keratitis<sub>(9, 0)</sub>, <sub>(10, 1)</sub> corneal ulcer<sub>(11, 2)</sub> etc<sub>(12, 3)</sub>, <sub>(13, 4)</sub>.”

where the subscript of the word represents the position information of the word in the sentence, the first figure represents the relative distance between the first entity and the word, and the second figure represents the relative distance between the second entity and the word. For example, “XX” represents the first entity, so the first figure in the subscript is 0, and “fungal keratitis” represents the second entity, so the relative distance between “XX” and the word is 9. In addition, as “XX” is before “fungal keratitis”, the second figure in the subscript of “XX” is -9. Thus, the position vectors are:

{[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13],

[-9, -8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4]}

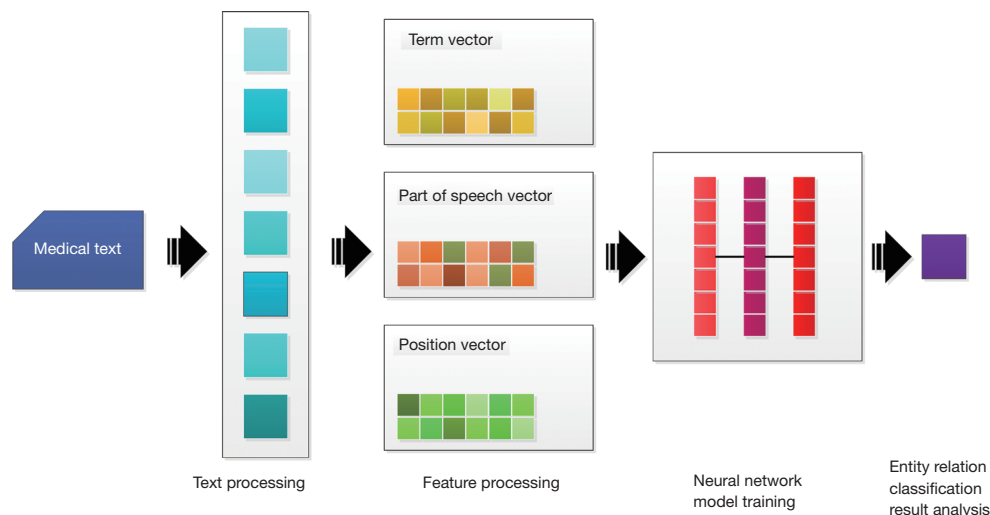
The first vector refers to the relative position vector of the first entity to the words, while the second vector refers to the relative position vector of the second entity to other words. The stacking of the 2 vectors forms the traditional entity position vector.

As the position of the subject term omitted is fixed, the position information of the subject term does not enable the model to judge entity relations. If the sentence were to be completed by adding the missing subject, the model would make unnecessary calculations during the training. Thus, in this study, the position information was represented by the position vector as follows:

“Used for<sub>(-8)</sub> eye<sub>(-7)</sub> infections<sub>(-6)</sub> caused by<sub>(-5)</sub> sensitive<sub>(-4)</sub> fungi<sub>(-3)</sub>, <sub>(-2)</sub> such as<sub>(-1)</sub> fungal keratitis<sub>(0)</sub>, <sub>(1)</sub> corneal ulcer<sub>(2)</sub> etc.<sub>(3)</sub>, <sub>(4)</sub>.”

The new position vector form was as follows:

[-8, -7, -6, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4].



**Figure 1** Data processing flow of the model.

### *Multi-scale CRNN entity relation extraction of a single entity*

The working process of the multi-scale CRNN (see *Figure 1*) was as follows:

- (I) Text processing: for medical texts, first segment the words, and then count up how many times each word appears in the text. The words that appear repeatedly but are meaningless are treated as “stop words.” Stop words can have a negative effect on model construction, and thus need to be removed.
- (II) Feature processing: generate the term vector, the part of the speech vector, and the position of the vector in with the data after text processing.
- (III) Neural network model training: train the neural network model with the above-mentioned 3 feature vectors.
- (IV) Entity relation classification result: obtain the entity relation classification by analyzing the results generated by the model.

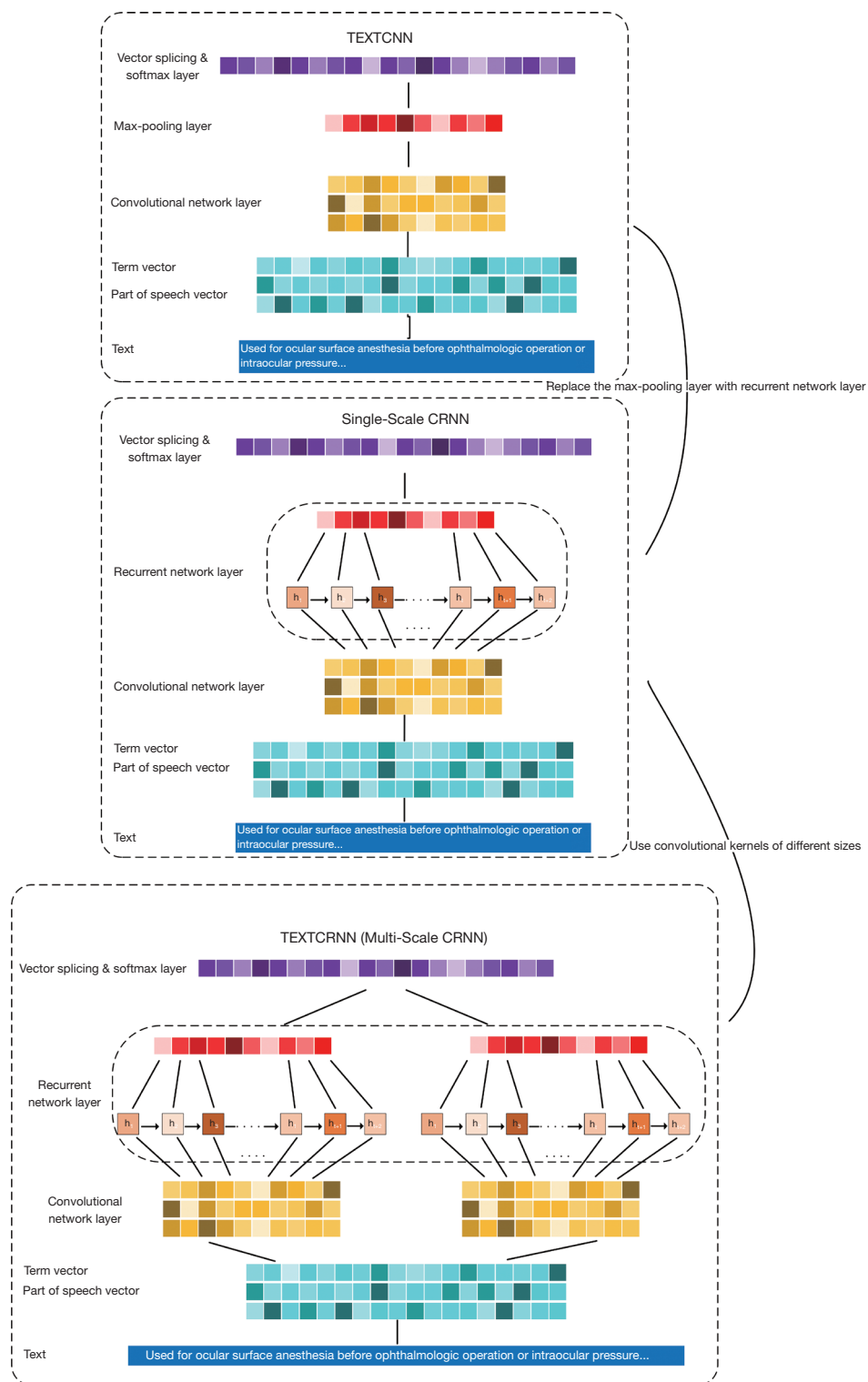
In this study, the neural network model used was the multi-scale CRNN, which combined the advantages of the CRNN and TEXTCNN. The CRNN captures deeper time-sequence characteristics from the text, while the TEXTCNN captures features at different scales from the text. By combining the 2 networks, the model was able to capture the time-sequence characteristics of the features at different scales from the text.

The text-oriented CRNN model was modified based

on the CNN model. In the CNN model, the CNN layer extracts the target vector features by convolution, and then uses the max-pooling layer to receive the feature vector output from the CNN layer (see the CNN in *Figure 2*). The function of the max-pooling layer in the network model is to refine the features obtained from the CNN layer and remove the insignificant features. As per the process described above, the model cannot capture the time-sequence characteristics of the target vector, but this problem can be solved with the CRNN model.

In the CRNN model, the RNN layer receives the feature vector output from the CNN layer (see the CRNN in *Figure 2*). The target vector output from the CNN layer are the local features that are relatively independent each other. When these vectors are input into the RNN layer, the time-sequence characteristics among the local features can be captured. In this way, the model not only captures the local features, but also captures the deeper time-sequence characteristics.

In the traditional CRNN, only 1 size of convolution kernels is used, but in the multi-scale CRNN, convolution kernels of several sizes are used to capture the local features at different scales (see the multi-scale CRNN in *Figure 2*). Additionally, if different sizes of convolution kernels are used in the CNN layer, the multi-scale CRNN can capture the time-sequence characteristics of features at different scales. Finally, the features captured by splicing the vectors in the model are integrated, and the final result is produced with SoftMax.



**Figure 2** Evolution process from the TEXTCNN model structure to the multi-scale CRNN model structure. CNN, convolutional neural networks; TEXTCNN, convolutional neural networks for sentence classification; Single-Scale CRNN, Single-Scale convolutional recurrent neural network for Sentence Classification; TEXTCRNN (Multi-Scale CRNN), multi-scale convolutional recurrent neural network for sentence classification.



**Table 1** Number of each entity relation

Type name	Quantity
The drug can be used for a disease	640
The drug can be used with another drug	565
The drug can be used for a symptom	536
The drug cannot be used for a disease	431
The drug should be cautiously used for a disease	216
The drug should be cautiously used for a group of people	181
The drug might cause a symptom	115
The drug might cause a disease	78
Total	2,762

### Statistical analysis

All Statistical analysis of medical literature data was performed using SPSS software (version 22.0, IBM Corp., Armonk, NY, USA). The quantitative data are expressed as the mean  $\pm$  standard deviation (SD) for normally distributed data, or the median (interquartile range) for non-parametric data. Categorical variables of 8 valid entity relations were analyzed using the chi-square tests and Student's *t*-test. *P* value  $<0.05$  was considered to be statistically significant.

## Results

### Data set

The experimental data set used included “indication”, “contraindication”, “adverse reactions” and “precautions”, which were captured in every chapter of the *Pharmacopoeia*. The data size of every chapter was still huge; thus, 3,800 sentences were randomly selected for entity relation labeling.

Among the 3,800 sentences, there were 2,762 pairs of valid entity relations, which were classified into the following 8 types (see *Table 1*):

- (I) The drug can be used for a disease: what disease can be treated with the drug?
- (II) The drug can be used for a symptom: what symptom can be treated with the drug?
- (III) The drug can be used with another drug: when the drug is used for treatment, what drug can be used with it?
- (IV) The drug cannot be used for a disease: the drug is

contraindicated in the disease. What disease does a patient suffer from when the drug cannot be taken?

- (V) The drug should be cautiously used for a disease: what disease does a patient suffer from when the drug should be used with caution, as adverse reactions may be caused after the drug is taken?
- (VI) The drug should be cautiously used for a group of people: what group of people should avoid the drug, as for some groups of people, the drug may cause adverse reactions?
- (VII) The drug may cause a disease: what disease might a patient suffer from after taking the drug?
- (VIII) The drug may cause a symptom: what symptom might a patient suffer from after taking the drug?

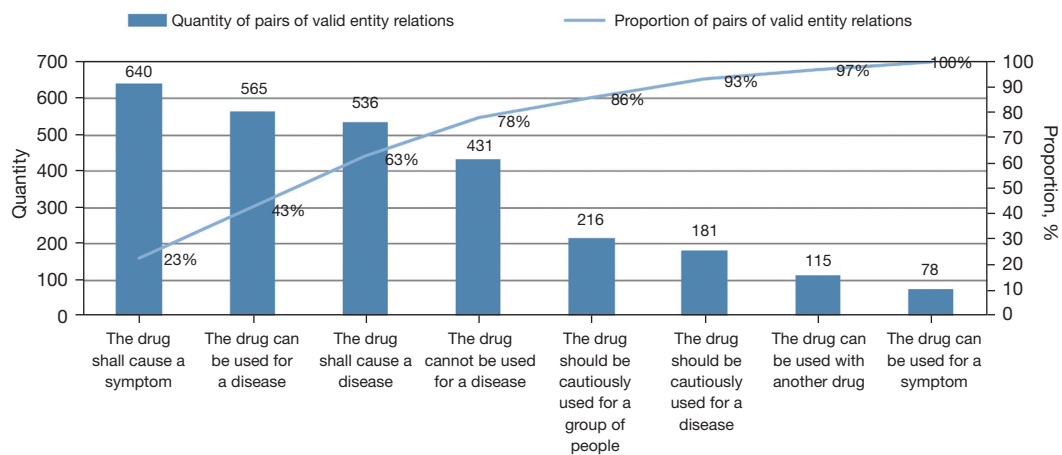
*Table 1* represents the number of each entity relation.

*Figure 3* is the Pareto chart for the quantities of entity relations. As the Pareto chart shows, the data size of the entity relation types decreases successively from left to right. The curve shows the accumulative trend of the entity relation types, and it is evident that the slope of the curve on the left side is significantly steeper than that on the right side. The reason is that the quantity of the entity relation types on the left side is far more than that on the right side.

When part of the speech is tagged in the data set, in addition to tagging the general part of the speech used, the words for which the entity relation extraction will be done in the text are also classified in greater detail. As the entities in the entity relation extraction are nouns, the proper nouns under the part of the speech classified as nouns are subdivided into “nouns—symptoms”, “nouns—diseases”, and “nouns—groups”, and thus the model is better able to capture the relation classification information.

### Part of speech definition

In data enhancement based on reinforcement learning, the probability of using the data enhancement method is calculated based on different parts of speech; thus, the definition of parts of speech is of great significance. In this study, we adopted the criteria for parts of speech of the jieba word segmentation tool. These criteria are based on the criteria for parts of speech developed by Beijing University, but some parts of speech were further subdivided and extended to conduct a more detailed classification of parts of speech. For example, in relation to common adjectives, the adjective phrase (“al”) is newly added into the criteria for the parts of speech in the jieba word segmentation. As



**Figure 3** Pareto chart for the quantities of entity relations.

the objects in the entity relation extraction from medical literature are usually proper nouns in the professional field, and the entity type of the entity pair is a significant entity relation feature, 4 types of nominal part-of-speech tagging were newly added based on the jieba criteria for parts of speech, which corresponded to “symptom”, “disease”, “drug” and “group” and that had relations to the drug. Thus, 73 types of part-of-speech tagging were used in this study, including 39 types from the table of Beijing University, 30 newly added types from the jieba criteria for parts of speech, and 4 types of proper nouns newly added in this study.

### Assessment standards

In this experiment, the model performance was analyzed from the following 2 perspectives: (I) the capability of the model for each entity relation type; and (II) the overall classification performance of the model with multiple types. The former was measured by precision, recall, and F score, while the latter was assessed by macro precision, macro recall, and macro-F score.

The assessment values above were calculated based on the prediction results. “True” or “False” represents whether the sample was correctly predicted, while “Positive” or “Negative” represents the sample’s prediction results; that is, whether the sample was predicted as a positive sample or a negative sample. There are 4 possible combinations, including true positive (TP), false positive (FP), true negative (TN), and false negative (FN).

Precision (also called the precision ratio), refers to the number of true predictions among all positive samples

predicted, and is calculated as:

$$Precision = \frac{TP}{TP + FP} \quad [1]$$

Accuracy refers to the number of true predictions among all the samples, and is calculated as:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad [2]$$

Recall (also called the recall ratio), refers to the number of true predictions among the actual positive samples, and is calculated as.

$$Recall = \frac{TP}{TP + FN} \quad [3]$$

Generally, a model cannot be simply measured by precision and recall. Precision and recall are always expressed as a negative correlation in the pulling and pushing state. In addition, it is not intuitive to compare 2 ratios among several models; thus, a new standard had to be introduced to combine the 2 ratios and intuitively assess the advantages and disadvantages of the model. The new standard is  $F_\beta$ -score, which is calculated as:

$$F_\beta = \frac{(1 + \beta^2) Precision \cdot Recall}{\beta^2 \cdot Precision + Recall} \quad [4]$$

where  $\beta$  represents the different weight of the precision ratio and the recall ratio in the combination, and  $F_\beta$  represents the importance degree of the recall ratio and precision ratio. Generally, the  $\beta$  is 1, and  $F_1$  represents the same importance of the recall ratio and the precision ratio.

**Table 2** Experiment model parameter setting

Parameter	Value
CNN length	4 and 6
The number of CNNs with the length of 4	16
The number of CNNs with the length of 6	16
Double-layer GRU	50
BiLSTM	50
Learn rate	0.01
Epoch	50

CNN, convolutional neural networks; GRU, gated recurrent unit; BiLSTM, bidirectional long- and short-term memory.

For the model with multiple types, the  $F_1$  values of the types cannot indicate the model performance. The reason is that each type may have different results for different models, and thus it is difficult to assess model performance. For this purpose, macro precision, macro recall and macro-F value were used for the performance assessment of the multi-classification model, which is calculated as:

$$precision_{ma} = \frac{\sum_{i=0}^I precision_i}{I} \quad [5]$$

$$recall_{ma} = \frac{\sum_{i=0}^I recall_i}{I} \quad [6]$$

$$macro F_1 score = 2 \frac{precision_{ma} \times recall_{ma}}{precision_{ma} + recall_{ma}} \quad [7]$$

### Experiment setting

In the experiment, the reference models included the original TEXTCNN and 2 other classical neural network structures of long short-term memory (LSTM) (21) and GRU (22). The BiLSTM model was adopted for the LSTM, and the double-layer stacking GRU was adopted for the GRU. As for the RNN part in the TEXTCRNN, BiLSTM and GRU were used separately to establish 2 models for a contrast experiment (see Table 2).

### Experiment results

Next, the parts of speech mentioned above were used

to tag the data set, and the classical network model and the TEXTCRNN were then separately used to train the entity relation extraction model for the text tagged. The experiment results are shown in Tables 3,4. In these tables, the highest F1 value in all types is highlighted with bold and underline.

Table 3 shows the assessment value of each entity relation type in the classical neural network models, which include the TEXTCNN, BiLSTM, and double-layer stacking GRU. Notably, among the 5 models used in the experiment, the highest F1 value of the 3 entity relation types was obtained in the TEXTCNN. Among the classical neural network models, the highest F1 value of the 4 entity relation types was obtained in the TEXTCNN, the highest F1 value of 1 entity relation type was obtained in the BiLSTM, and the highest F1 value of 3 entity relation types was obtained in the double-layer stacking GRU. If only the assessment values of the entity relation types are considered, there was no significant difference in the performance of the 3 classical models.

Table 4 shows the TEXTCRNN neural network models with different RNN structures. Notably, it is evident that the highest F1 value of the 3 entity relation types was obtained for the 2 TEXTCRNN models. Additionally, if only the 2 network models are compared, the highest F1 value of the 4 entity relation types was obtained for each of them.

When the 2 tables are combined, among all types, the number of best experimental results from the TEXTCNN, TEXTCRNN(BiLSTM), and TEXTCRNN (double-layer stacking GRU) was 3, 2 and 3, respectively; thus, no outstanding achievement is obtained by purely using the neural network with the RNN structure. Additionally, it should be noted that the F1 values of the different relation types differed greatly. The F1 value of 1 relation type was up to 0.91, but that of another was only 0.61. For most types, the new position information coding method can effectively provide the basis for entity relation classification.

Table 5 shows the micro indicators of each model.

### Discussion

In entity relation extraction for a single entity, compared to the other models, the TEXTCRNN (double-layer stacking GRU) had a better capability of entity relation classification when the same entity word did not belong to the same entity relation. For example, consider the phrase “high blood pressure” in the following sentences: “high blood pressure is mainly mild to moderate, and always appears



**Table 3** Assessment value of each type in the classical neural network models

Category	TEXTCNN			BiLSTM			Double-layer stacking GRU		
	Precision	Recall	F1 value	Precision	Recall	F1 value	Precision	Recall	F1 value
The drug will cause a symptom	0.6824	0.8016	0.7372	0.6098	0.7937	0.6897	0.6797	0.7768	0.7250
The drug can be used for a disease	0.7593	0.7664	0.7628	0.8132	0.6916	0.7475	0.7895	0.8257	0.8072
The drug will cause a disease	0.7340	0.5308	0.6161	0.6559	0.4692	0.5471	0.6970	0.6161	0.6540
The drug cannot be used for a disease	0.9048	0.8837	0.8941	0.8046	0.8140	0.8092	0.8953	0.9625	0.9277
The drug should be cautiously used for a group of people	0.8250	0.9429	0.8800	0.8000	0.9143	0.8533	0.8500	0.7727	0.8095
The drug should be cautiously used for a disease	0.7500	0.8824	0.8108	0.6750	0.7941	0.7297	0.9348	0.8600	0.8958
The drug can be used with another drug	0.5000	0.6471	0.5641	0.5238	0.6471	0.5789	0.5200	0.5652	0.5417
The drug can be used for a symptom	0.8235	0.7778	0.8000	0.6471	0.6111	0.6286	0.9333	0.6087	0.7368

TEXTCNN, convolutional neural networks for sentence classification; GRU, gated recurrent unit; BiLSTM, bidirectional long- and short-term memory.

**Table 4** The assessment value of each type in the TEXTCRNN with different RNN structures

Assessment standard	Model					
	TEXTCRNN (BiLSTM)			TEXTCRNN (double-layer stacking GRU)		
	Precision	Recall	F1 value	Precision	Recall	F1 value
The drug will cause a symptom	0.7345	0.6587	0.6946	0.6779	0.8016	0.7345
The drug can be used for a disease	0.8286	0.8131	0.8208	0.8056	0.8131	0.8093
The drug will cause a disease	0.6312	0.6846	0.6568	0.7000	0.5385	0.6087
The drug cannot be used for a disease	0.9024	0.8605	0.8810	0.8370	0.8953	0.8652
The drug should be cautiously used for a group of people	0.8649	0.9143	0.8889	0.8919	0.9429	0.9167
The drug should be cautiously used for a disease	0.8158	0.9118	0.8611	0.9375	0.8824	0.9091
The drug can be used with another drug	0.5789	0.6471	0.6111	0.5556	0.5882	0.5714
The drug can be used for a symptom	0.7778	0.7778	0.7778	0.8235	0.7778	0.8000

TEXTCRNN, multi-scale convolutional recurrent neural network for Sentence Classification; GRU, gated recurrent unit; BiLSTM, bidirectional long- and short-term memory; RNN, recurrent neural networks.

in the early stage after the patient takes the drug. It can be controlled with common hypotensive drugs”, and “high blood pressure, but cannot be used as a first-line drug, and usually used as a second-line or third-line treatment drug to be used together with other hypotensive drugs”. The 2 sentences belong to different entity relations. The former is about the symptom appearing after the drug is taken, and thus belongs to “the drug causes a symptom” type, and the later is about the disease that can be treated with the drug, and thus belongs to “the drug can be used for a

disease” type. *Table 6* sets out prediction results based on the 2 above sentences in each model. The prediction results show that only the TEXTCRNN (double-layer stacking GRU) correctly predicted to which entity relation type “high blood pressure” belonged, and in the other models, the entity relation to which that the entity belonged was misjudged.

The prediction difference among the models above may have occurred for the reasons outlined below.

(I) Position information. Due to the lack of subject

**Table 5** Micro indicators of each model

Model	MARCO PRECISION	MARCO RECALL	MARCO F1_SCORE
TEXTCNN	0.747378	0.779060	0.762890
BiLSTM	0.691165	0.716875	0.703785
GRU	0.787449	0.748461	0.767460
TEXTCRNN (BiLSTM)	0.766764	0.783473	0.775028
TEXTCRNN (Double-layer stacking GRU)	0.778605	0.779963	0.779284

Notably, the model with TEXTCRNN (double-layer stacking GRU) was superior to other classical models. TEXTCNN, convolutional neural networks for sentence classification; TEXTCRNN, multi-scale convolutional recurrent neural network for Sentence Classification; GRU, gated recurrent unit; BiLSTM, bidirectional long- and short-term memory.

**Table 6** Prediction results

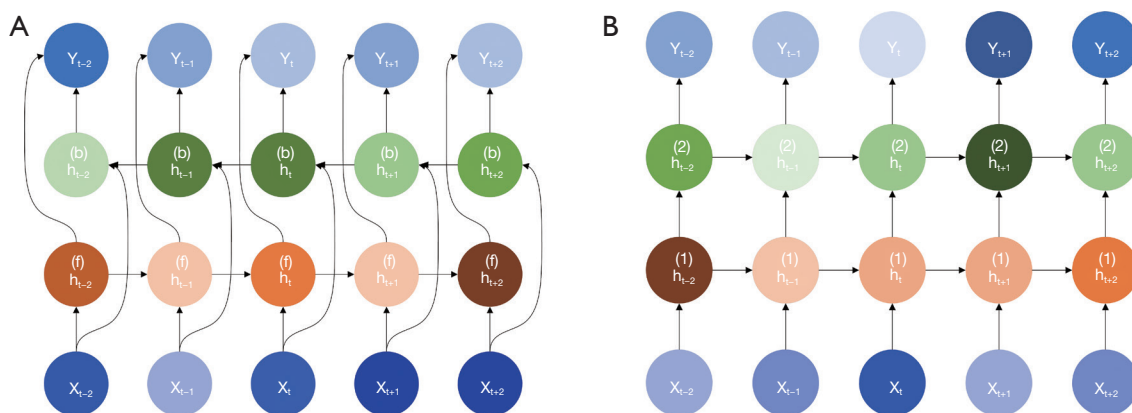
Sentence	Actual type	TEXTCNN, predicted type	GRU, predicted type	BiLSTM, predicted type	TEXTCRNN (double-layer stacking GRU), predicted type	TEXTCRNN (BiLSTM), predicted type
High blood pressure is mainly mild to moderate, and always appears in the early stage after the patient takes the drug. It can be controlled with common hypotensive drugs.	The drug will cause a symptom	The drug will cause a symptom	The drug will cause a disease	The drug will cause a disease	The drug will cause a symptom	The drug will cause a disease
High blood pressure, but cannot be used as a first-line drug, and usually used as a second-line or third-line treatment drug to be used together with other hypotensive drugs.	The drug can be used for a disease	The drug can be used with another drug	The drug can be used with another drug	The drug can be used with another drug	The drug can be used for a disease	The drug will cause a disease

TEXTCNN, convolutional neural networks for sentence classification; TEXTCRNN, multi-scale convolutional recurrent neural network for Sentence Classification; GRU, gated recurrent unit; BiLSTM, bidirectional long- and short-term memory.

information, the models could not learn the relative position information between entities; thus, they could only learn the entity relation features from the position information of a single entity. In the 2 sentences used for prediction in *Table 6*, in addition to there being no entity term that can be associated with “high blood pressure,” the term “high blood pressure” also appears at the beginning of the sentences. Thus, the position vectors of the 2 sentences are the same; thus, only the contextual information could be used to judge the entity relation type to which the entity term belongs, but the contextual information is not sufficient to correctly distinguish the entity relation type.

- (II) Network structure. The difference in network structure may also be a reason resulting for the

difference in the results. *Figure 4* shows the diagrams of 2 neural network structures. Notably, the same vector was input into the 2 layers of the neural network of the BiLSTM, and their difference was that the vector direction in the 2 layers was opposite; however, in the double-layer stacking GRU, the output vector from the previous layer of neural network was the input vector into the subsequent layer of neural network, and their direction was the same. Such structure difference indicates that the BiLSTM can only captured the time-sequence characteristics of the same input vector from 2 directions, but the double-layer stacking GRU also captured deeper time-sequence characteristics based on the time-sequence characteristics that had been captured from the input vector in the first layer. It may



**Figure 4** Two neural network structures. (A) BiLSTM neural network layer structure. (B) Double-layer stacking GRU neural network structure. BiLSTM, bidirectional long- and short-term memory; GRU, gated recurrent unit.

also be a reason why the TEXTCRNN (double-layer stacking GRU) could accurately distinguish the entity relation based only on the contextual information.

The convolution operations of the CNN were relatively independent, so other than the range covered by the current convolution kernel, the other terms were not correlated. To increase the connections among the terms, the size of the convolution kernel needs to be enlarged; however, too large a convolution kernel will increase the model parameters, and thus increase the difficulty of model training. To solve the long-term dependency problem, we sought to construct the model network by replacing the max-pooling layer with the RNN layer, as in the RNN, the neurons are connected end to end, so that information generated at different times are transmitted among neurons, and the terms with relatively long distances between them are associated. Additionally, when the convolution kernels of the CNN were applied in NLP, they had the same advantages as n-gram; thus, different convolution kernels can be used to capture text features at different scales. Based on the experimental results of Sahu *et al.* (20), the convolution kernels with the length of 4 and 6 were used for initial feature capture from the text.

As shown in the experiment, compared to the classical neural networks, such as the TEXTCNN, GRU, and BiLSTM, the neural network with the TEXTCRNN structure was significantly improved by integrating the network features of CNN and RNN as well as adopting 2 scales of convolution kernels. Further, the new position information coding method effectively provided a basis for the entity relation classification. However, there are some

deficiencies in the experiment; for example, the attention mechanism was not applied, and the pre-trained term vector was not applied. The attention mechanism was not used to improve the model effectiveness by highlighting the significance of different positions, as this study aimed only to compare the network structures and reduce the effect of other technologies on network structures as far as possible, regardless of whether the affect was positive or negative. The reason why the pre-trained term vector was not adopted was that compared to the general data set, medical literature include many proper nouns, which trigger the out-of-vocabulary problem; thus, the method of training our own term vectors was adopted in this study. However, as the text used to train our own term vectors was limited, the text features captured by the term vectors were limited. In the follow-up research, medical literature and textbooks will be used as text corpus to train word vectors, so as to further improve the features of word vectors.

The biggest problem facing the methodology of this study is that in ordinary relationship extraction, there are usually two entities in the text at the same time, while the sentences in the Pharmacopoeia used in this paper generally have only one entity. Therefore, it is necessary to solidify the subject entity at the beginning of the sentence and forcibly establish the relationship between the two entities, such treatment will limit the scope of use of the model in this paper.

## Acknowledgments

**Funding:** This work was supported by Mobile Health Ministry of Education, China Mobile Joint Laboratory

Project, Research and Application of DRGs Grouping System Based on Big data (No. 2020MHL02015).

### Footnote

*Data Sharing Statement:* Available at <https://atm.amegroups.com/article/view/10.21037/atm-22-1226/dss>

*Conflicts of Interest:* All authors have completed the ICMJE uniform disclosure form (available at <https://atm.amegroups.com/article/view/10.21037/atm-22-1226/coif>). XW, LL, JL are from Hunan Creator Information Technology Co. Ltd. The other authors have no conflicts of interest to declare.

*Ethical Statement:* The authors are accountable for all aspects of the work, including ensuring that any questions related to the accuracy or integrity of any part of the work have been appropriately investigated and resolved.

*Open Access Statement:* This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

### References

- Gupta A, Banerjee I, Rubin DL. Automatic information extraction from unstructured mammography reports using distributed semantics. *J Biomed Inform* 2018;78:78-86.
- Topaz M, Murga L, Gaddis KM, et al. Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches. *J Biomed Inform* 2019;90:103103.
- Lee KH, Kim HJ, Kim YJ, et al. Extracting Structured Genotype Information from Free-Text HLA Reports Using a Rule-Based Approach. *J Korean Med Sci* 2020;35:e78.
- Srivastava P, Bej S, Yordanova K, et al. Self-Attention-Based Models for the Extraction of Molecular Interactions from Biological Texts. *Biomolecules* 2021;11:1591.
- Armean IM, Lilley KS, Trotter MWB, et al. Co-complex protein membership evaluation using Maximum Entropy on GO ontology and InterPro annotation. *Bioinformatics* 2018;34:1884-92.
- Jalal A, Khalid N, Kim K. Automatic Recognition of Human Interaction via Hybrid Descriptors and Maximum Entropy Markov Model Using Depth Sensors. *Entropy (Basel)* 2020;22:817.
- Chauhan VK, Dahiya K, Sharma A. Problem formulations and solvers in linear SVM: a review. *Artificial Intelligence Review* 2019;52:803-55.
- An N, Xiao Y, Yuan J, et al. Extracting causal relations from the literature with word vector mapping. *Comput Biol Med* 2019;115:103524.
- Sargül M, Ozyildirim BM, Avci M. Differential convolutional neural network. *Neural Netw* 2019;116:279-87.
- Yamashita R, Nishio M, Do RKG, et al. Convolutional neural networks: an overview and application in radiology. *Insights Imaging* 2018;9:611-29.
- Liimatainen K, Huttunen R, Latonen L, et al. Convolutional Neural Network-Based Artificial Intelligence for Classification of Protein Localization Patterns. *Biomolecules* 2021;11:264.
- Lerner I, Paris N, Tannier X. Terminologies augmented recurrent neural network model for clinical named entity recognition. *J Biomed Inform* 2020;102:103356.
- Alt C, Hübner M, Hennig L. Fine-tuning pre-trained transformer language models to distantly supervised relation extraction. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 2019;1388-98.
- Niu C, Zhao X. Study on the Method of Extracting Diabetes History from Unstructured Chinese Electronic Medical Record. *Parallel Architectures Algorithms and Programming* 2020:140-6.
- Hassan A, Mahmood A. Convolutional recurrent deep learning model for sentence classification. *IEEE Access* 2018;6:13949-57.
- Zaman MMA, Mishu SZ. Convolutional recurrent neural network for question answering. In *2017 3rd International Conference on Electrical Information and Communication Technology (EICT)* 2017;1-6.
- Raj D, Sahu S, Anand A. Learning local and global contexts using a convolutional recurrent network model for relation classification in biomedical text. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, 2017;311-321.
- Li Z, Yang Z, Shen C, et al. Integrating shortest dependency path and sentence sequence into a deep learning framework for relation extraction in clinical text. *BMC Med Inform Decis Mak* 2019;19:22.

19. Kim Y. Convolutional neural networks for sentence classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014;1746-1751.
20. Sahu S, Anand A, Oruganty K, et al. Relation extraction from clinical texts using domain invariant convolutional neural network. In Proceedings of the 15th Workshop on Biomedical Natural Language Processing, 2016;206-215.
21. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735-80.
22. Cho K, Van Merriënboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Computer Science* 2014. doi: 10.3115/v1/D14-1179.

**Cite this article as:** Chen T, Wu X, Li L, Li J, Feng S. Extraction of entity relations from Chinese medical literature based on multi-scale CRNN. *Ann Transl Med* 2022;10(9):520. doi: 10.21037/atm-22-1226