

A Named Entity Recognition Method Enhanced with Lexicon Information and Text Local Feature

Yuekun MA, He LIU, Dezheng ZHANG*, Chang GAO, Yujue LIU

Abstract: At present, Named Entity Recognition (NER) is one of the fundamental tasks for extracting knowledge from traditional Chinese medicine (TCM) texts. The variability of the length of TCM entities and the characteristics of the language of TCM texts lead to ambiguity of TCM entity boundaries. In addition, better extracting and exploiting local features of text can improve the accuracy of named entity recognition. In this paper, we proposed a TCM NER model with lexicon information and text local feature enhancement of text. In this model, a lexicon is introduced to encode the characters in the text to obtain the context-sensitive global semantic representation of the text. The convolutional neural network (CNN) and gate joined collaborative attention network are used to form a text local feature extraction module to capture the important semantic features of local text. Experiments were conducted on two TCM domain datasets and the *F1* values are 91.13% and 90.21% respectively.

Keywords: attention mechanism; char-word fusion coding; gate mechanism; NER

1 INTRODUCTION

Traditional Chinese medicine (TCM) is one of the common means of modern medical treatment. With the continuous development of medical technology, TCM needs to continuously improve its knowledge system. Through thousands of years of inheritance, TCM has accumulated a large amount of valuable experience and literature, containing a wealth of scientific knowledge of TCM. TCM ancient literature is the crystallization of the wisdom of generations of TCM practitioners, containing clinical knowledge of TCM such as diagnosis, prescription and medication rules, which has important practical significance to the development of modern clinical treatment. In recent years, how to extract useful TCM knowledge from unstructured ancient TCM texts and apply it to practical life has received a lot of attention from researchers.

Named Entity Recognition (NER) [1] as a method of knowledge acquisition refers to identifying entities with special meaning from unstructured texts and classifying them into predefined meaningful categories. Generally, the TCM named entity recognition task is defined as a sequence labelling problem, aiming at assigning labels to each entity in the input text sequence. The variability of the length of TCM entities and the characteristics of the language of TCM texts result in the entity boundary being blurred. For example, "abdominal pain and bloating" is a symptom entity, including "abdominal pain". This problem makes many named entity recognition models designed for English corpus ineffective in the dataset of ancient Chinese medicine books. In view of the unclear boundary of Chinese medicine entities, the existing Chinese medicine named entity recognition methods mainly use combined characters and word embedding as the input of the model [2, 3] to help the model more accurately model word features. In these methods, the word acquisition method is the Chinese word segmentation tool, while the text language of ancient Chinese medicine books is mostly ancient Chinese, which has a large gap with modern Chinese, leading to the occurrence of incorrect word segmentation, and thus affecting the performance of the model. In addition, the method based on word

segmentation cannot get all the potential words of the sentence, which will also affect the accuracy of the prediction results of the model in practical applications. Therefore, accurately capturing all potential words in ancient Chinese medicine texts is the key to improve the performance of the model.

In addition, at the model construction level, most of the existing methods focus on extracting the context semantic features of text. For example, Zhang et al. [4] and Deng et al. [5] use BiLSTM to capture the context semantic features. Ma et al. [6] designed a multi-granularity text encoder to extract the context semantic features of text from multiple dimensions. The above work focuses on the extraction of the overall contextual information of the text. By analyzing the language characteristics of ancient Chinese medicine texts, we find that some Chinese medicine knowledge expressions are presented in the form of phrases, and there is no complete grammatical structure. Therefore, when building the model, we should consider the context semantic features and local semantic features of the input text at the same time, and make full use of their advantages to improve the overall performance of the model. The lexicon [7] is used to enhance the entity boundary discrimination of the model, and the local features of the text are used to enhance the semantic discrimination of the model.

In the task of named entity recognition, feature extraction of text is very important. The effectiveness and richness of feature information often determines the effect of entity recognition model. In recent years, many studies have shown that attention mechanism [8, 9] and gate mechanism [10, 11] have achieved ideal results in feature extraction and screening. Attention mechanism can usually distinguish the degree of information being focused by weight to determine the characteristics of information being focused. Therefore, we use attention mechanism to fuse character information and word information. The gate mechanism controls the flow of information in the model through some weighting functions, and can adaptively select and reject information according to the importance of feature information. Therefore, we apply gate mechanism to text feature fusion and local feature enhancement.

Based on the above analysis, we propose a TCM named entity recognition model with lexicon information and text local feature enhancement. Specifically, at the model input level, we will match the input text with a Chinese medicine lexicon to obtain all potential word information, and generate corresponding word sequences for the input text sequence. In order to efficiently use all the matched word information, we use the attention mechanism to dynamically extract the most relevant word semantic information for each character. In the aspect of local feature modelling, we propose a local feature awareness network, which is composed of two parts: a dual-channel convolutional neural network (CNN) unit, which can capture local semantic features of different granularity; a gate joined collaborative attention network, which can use local information of different granularity to achieve mutual reinforcement. In addition, the dual-channel CNN unit can also capture the potential semantic features of words when capturing local features, and can compensate for the impact of unlisted words in the lexicon on the performance of the model. Additionally, we designed complementary fusion gates to use to adaptively fuse contextual semantic features with local semantic features. The Puji Fang and the Materia Medica are authoritative works on TCM prescriptions and Chinese herbal medicine, and we used them as experimental data. The main contributions of this paper are as follows.

(1) To obtain better performance of the model on the TCM domain corpus, we created a TCM lexicon containing a total of 102 800 words related to prescription names, herb names, symptom names, tongue names, pulse names, evidence names and other related words, and generated word embedding vector representations for each word using a self-attentive mechanism and positional features.

(2) We proposed a two-branch model architecture, which jointly utilizes a fused lexicon fusion approach and a local feature extraction approach for the NER task.

(3) Experiments on two TCM data sets show that our method improves 0.51% and 0.73% compared with the optimal baseline. The ablation experiments and analysis also justified well the method in this paper.

2 METHODS

In this section, we first introduced the construction of the lexicon and the construction method of the word vector, and then illustrated the architecture of the TCM named entity model based on multi-feature fusion, and gave an example to show how our model works.

2.1 TCM Lexicon and Word Vector Construction

The vocabulary in the TCM lexicon constructed in this paper is derived from the TCM-related words in Sogou's thesaurus, TCM medical cases, Chinese herbal medicine manual and other related documents, and a total of 102800 TCM words are collected. Since traditional word separation tools have poor effect on the word separation of ancient TCM texts, the word embedding vector representation of the vocabulary cannot be generated directly using the large-scale corpus. Therefore, we first

used Word2vec [12] to convert the characters in each vocabulary into a dimensional embedding vector representation, which are trained by relying on the large-scale TCM text corpus, and connect these embedding vectors to form a feature matrix representation of the vocabulary as follows.

$$word_embedding = word2vec(Word) \quad (1)$$

where $Word$ is the vocabulary to be vectorized and $word_embedding$ is the feature matrix after vectorization. Then, we consider the positional features of each character in the vocabulary, so the positional feature vector p is added to each character embedding representation to obtain the feature matrix. The positional features of the t -th character are calculated as:

$$\begin{aligned} PE_{t,2i} &= \sin(t / 10000^{2i/d}) \\ PE_{t,2i+1} &= \cos(t / 10000^{2i/d}) \end{aligned} \quad (2)$$

where i takes a range of values $\left[0, \frac{d}{2}\right]$ and d denotes the dimensionality of the input character vector. Finally, we encode the feature matrix using the self-attentive mechanism [13] and obtain the final word embedding vector representation $word_vec$ using the average pooling method as follows.

$$word_vec = mean_pooling(SelfAttention(L)) \quad (3)$$

where $SelfAttention$ denotes the self-attentive mechanism and $mean_pooling$ denotes the average pooling.

2.2 Model Framework

The overall framework of our proposed model is shown in Fig. 1, which consists of four main components: encoder module, lexicon enhanced text global feature extraction module, text local feature extraction module and decoder module. The encoder module performs vectorized representation of the input character and word sequences. Lexicon enhanced text global feature extraction module serves to integrate words into the character information, aiming to use word information to enhance sentence semantic information and clarify word boundary information. Text local feature extraction module serves to capture potential word information and local features in the original semantic information, and the decoder module serves to adaptively fuse features and predict the class of entity labels.

2.2.1 Encoder Module

1) Word Sequence Construction

Chinese sentences are usually expressed as a sequence of characters and do not contain word information. This phenomenon makes the model unable to clearly model the word information, resulting in poor model effect. In order to make full use of the word information in the sentence, we use the lexicon to identify all potential words in the sentence and generate a new word sequence.

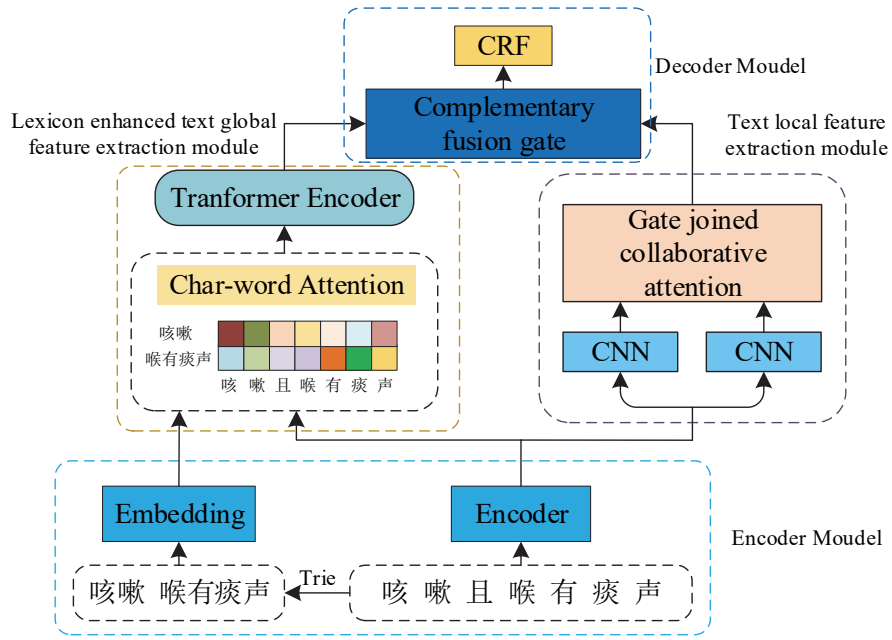


Figure 1 Model framework

Specifically, given a lexicon and a sentence, we build a lexicon tree for the lexicon, then traverse all characters in the sentence, and obtain all words by matching with the lexicon tree and form a word sequence in turn, as the example shown in Fig. 1, "哮喘且喉有痰声". After lexicon matching, we got a sequence of words {"哮喘", "喉有痰声"}.

2) Encoder and Embedding

We used a pre-trained model for vectorized representation of sentences, which addresses the drawbacks of traditional static word vector methods, and the pre-trained model contained rich prior knowledge, which is useful for NER tasks [14, 15]. Given a sentence input $X^c = \{x_1^c, \dots, x_i^c, \dots, x_n^c\}$, n denotes the length of the input text sequence. The sentence is represented as an embedding sequence $E^c = \{e_{CLS}^c, e_1^c, \dots, e_i^c, \dots, e_n^c\}$ using the pre-trained model, with e_i^c denoting the vector representation corresponding to each character in the sentence and e_{CLS}^c denoting the overall semantic information of the sentence. The encoding process of the sentence can be represented formally as:

$$E^c = Encoder(X^c) \tag{4}$$

where d^c of $E^c \in R^{n \times d^c}$ denotes the dimensionality of the hidden states output by the pre-trained model.

We used the word vector constructed above to convert the generated word sequence $X^w = \{x_1^w, \dots, x_i^w, \dots, x_m^w\}$ into a word embedding sequence $E^{w1} = \{e_1^{w1}, \dots, e_i^{w1}, \dots, e_m^{w1}\}$, where m denotes the length of the word sequence and d^{w1} of $E^{w1} \in R^{n \times d^{w1}}$ denotes the dimensionality of the word vector. The word sequence encoding process can be represented formally as:

$$E^{w1} = Embedding(X^w) \tag{5}$$

2.2.2 Lexicon Enhanced Text Global Feature Extraction Module

1) Char-word Attention (CW-ATT)

The input sentence gets the word sequence after lexicon matching, and our goal is to fuse the word information and character information. Inspired by the study of li et al. [16], we proposed the character-word attention mechanism. Unlike Li et al. we performed attentional computation on character sequences and word sequences directly, instead of self-attentional computation after splicing word sequences to character sequences, which has the advantage of reducing computational cost.

The character-word attention takes the character-embedding sequence E^c and the word-embedding sequence E^{w1} as input. Before calculating the attention weights, the dimensions of the two vectors need to be aligned, i.e., the word-embedding sequence $E^{w1} \in R^{n \times d^{w1}}$ dimensions are first transformed to obtain $E^{w2} \in R^{n \times d^{w2}}$, $d^{w2} = d^c$, and then the correlation between the character-embedding sequence E^c and the word-embedding sequence E^{w2} is calculated and the attention matrix is obtained as follows.

$$E^{w2} = W_2(\tanh(W_1 E^{w1} + b_1)) + b_2 \tag{6}$$

$$Att = \text{soft max}(E^c (E^{w2})^T) \tag{7}$$

where W_1 and W_2 are trainable weight matrices and b_1 and b_2 are bias matrices. Then we computed the weighting matrices of the character embedding sequence E^c and the attention matrix Att :

$$Z = E^c \cdot Att \tag{8}$$

Finally, we injected the weighted information into the character embedding sequence:

$$E^{c+w} = E^c + Z \quad (9)$$

Subsequently, the E^{c+w} was encoded using the discard and normalization layers.

2) Transformer Encoder

The character-word attention incorporates word information into character information, but the contexts are not sufficiently interacted with each other such that the current character embedding sequence E^{c+w} cannot express the full semantics of the sentence. We used the encoder part of Transformer to fully interact with the character-embedded sequence E^{c+w} incorporating word information. Transformer encoder consists of a multi-headed attention layer, a feed-forward neural network layer, and a layer normalization and residual network with good global modelling capability and parallel computing capability.

Given the character embedding sequence E^{c+w} , first added the position embedding information to E^{c+w} get the embedding sequence H , and then to input H into the Transformer encoder, the calculation formula is as follows.

$$\begin{aligned} G &= LN(H + MultiHAtt(H)) \\ H_G &= LN(G + FFN(G)) \end{aligned} \quad (10)$$

where *MultiHAtt* is the multi-headed attention mechanism, *LN* is the layer normalization, and *FFN* is the feed-forward neural network. The equation for multi-headed attention is as follows.

$$Att(K^{(h)}, Q^{(h)}, V^{(h)})^{(h)} = \text{soft max}\left(\frac{Q^{(h)}(K^{(h)})^T}{\sqrt{d_k}}\right)V^{(h)} \quad (11)$$

$$MultiHAtt(H) = [Att^{(1)}, \dots, Att^{(h)}] \quad (12)$$

where $K^{(h)}$, $Q^{(h)}$, $V^{(h)}$ are obtained by mapping the embedding sequence H , and h denotes the index of head.

2.2.3 Text Local Feature Extraction Module

There are many special cases in TCM ancient texts, for example, "牛黄解毒丸" and "牛黄" belong to the category of prescriptions and material medical respectively, so Local feature modelling plays a role in the task of TCM named entity recognition. CNN [17, 18] is widely used in image processing because of their efficient local modelling performance. Therefore, we used CNN to model the local features of the sentences. Local feature modelling can capture not only local semantic information but also potential word information, which can help the model to clarify the boundary information of entities.

The analysis of the linguistic features of TCM ancient texts shows that TCM knowledge entities usually consist of two-word words and four-word words, which is also argued in the literature [19]. From the starting point of the textual characteristics of TCM ancient texts, we designed a

two-channel CNN to capture different granularity local features by setting up one-dimensional CNNs with window sizes of 2 and 4, respectively, calculated as follows.

$$H^k = \tanh(\text{CNN}(E, W^k, b^k)) \quad (13)$$

where $H^k \in d^k$ one-dimensional CNN generates the local feature vector, k denotes the size of the convolution kernel, and W^k , b^k denote the learnable weight matrix and bias parameters, respectively.

The local text semantic information of interest to different entities is different. Therefore, enhancing the relevant local textual feature representation can identify entities more accurately. With this motivation, we proposed a gate joined collaborative attention network based on the collaborative attention network [20], which controls the information flow by interacting local features of different granularity and using a gate mechanism to achieve the purpose of enhancing local features while being able to preserve the semantic features of the underlying vocabulary. The structure of the gate joined collaborative attention network is shown in Fig. 4, and the specific implementation process is shown in Eqs. (14) to (18).

$$C = A^T B \quad (14)$$

$$A' = A \otimes C, B' = B \otimes C^T \quad (15)$$

$$G_A = \sigma(W^A[A, A']), G_B = \sigma(W^B[B, B']) \quad (16)$$

$$H_A = FF(A + A'G_A), H_B = FF(B + B'G_B) \quad (17)$$

$$H_L = W_{LF}([H_A, H_B]) + b_{LF} \quad (18)$$

where C denotes the similarity matrix, A' and B' are the outputs of the collaborative attention network, σ is the sigmoid activation function, FF denotes the feed-forward neural network, $[,]$ denotes the cascade operation, W^A and W^B are the trainable weight matrices, b_{LF} is the bias vector, and H_L denotes the local feature vector.

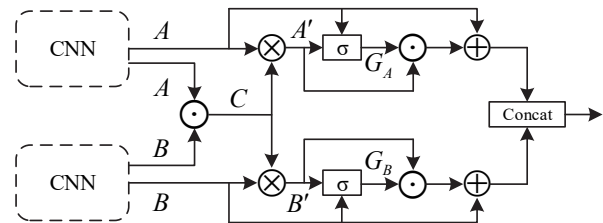


Figure 2 Gate joined collaborative attention mechanism

Specifically, Eq. (14) is used to calculate the similarity matrix between local features of different granularity, denoted as C , and then the interaction feature matrices, denoted as A' and B' , are computed separately using Eq. (15). The flow of information is controlled by gate σ in Eq. (16), which serves to retain the common semantic features related to the current semantic features while forgetting the

irrelevant common semantic features to avoid information redundancy, denoted as G_A and G_B . The feature vectors G_A and G_B are incorporated into the local feature vectors G_B and B , respectively, by Eq. (17), and then the final feature vector representations H_A and H_B are obtained by a feed-forward neural network. Finally, the local feature vector H_L is calculated using Eq. (18).

2.2.4 Decoder Module

1) Complementary Fusion Gate

In this section, we designed the complementary fusion gate to control the information flow and to be able to select the important information from the contextual semantic features and local semantic features. As shown in Fig. 1, the inputs of the complementary fusion gate are: the contextual semantic feature representation of the fused lexicon information and the local semantic feature representation of the unfused lexicon information. The output of the complementary fusion gate is represented as:

$$z = \sigma(W_z^1 H_G + W_z^2 H_L + b_z) \tag{19}$$

$$H = z \cdot H_G + (1 - z) \cdot H_L$$

where $\sigma(\cdot)$ is the sigmoid function, W_z^1 and W_z^2 are the learnable weight matrices, and b_z is the bias matrix. The feature matrices H_G and H_L have been transformed into the same dimension before the calculation of Eq. (19).

The complementary fusion gate aims to further enhance the representation of contextual semantic features of the information in the fused lexicon using latent word semantic features captured by the dual-channel local feature-aware network and local semantic features of different granularity. We believed that directly incorporating H_L into H_G or assigning weights only to H_L would result in information redundancy, so we suggested assigning weights to both H_G and H_L to achieve an adaptive balance between both.

2) Decoder

Conditional random field (CRF) [21] is able to consider the dependencies between successive labels, enhance the constraint information of the before and after labels, and can be learned automatically during model training. Therefore, this paper used CRF as a decoder to obtain the global optimal tag sequence. For a TCM sentence, its probability can be expressed as:

$$Score(x, y) = \sum_{i=1}^n (P_{i,y_i} + T_{y_i,y_{i+1}}) \tag{20}$$

where P_{i,y_i} denotes the predicted label probability of character x_i and T is the transfer matrix. The conditional probability of the actual output label sequence y is:

$$p(y | x) = \frac{e^{Score(x,y)}}{\sum_{y \in Y_x} e^{Score(x,y)}} \tag{21}$$

Finally, the optimal sequence [22] is calculated using the Viterbi algorithm [23] with an objective loss function as:

$$L = \sum_{i=1}^n \log(p(y | x)) + \lambda \|\theta\| \tag{22}$$

where λ is the hyper parameter of L2 regularization and θ denotes the trainable parameter.

3 EXPERIMENTAL SETTINGS

3.1 Datasets

To evaluate our model, we conducted experiments on two TCM datasets, the TCM prescription dataset and the TCM material medical dataset, respectively. The TCM prescription dataset was derived from the Puji Fang, and the TCM herbal dataset was derived from the Material Medical. The data were manually labelled under the guidance of TCM experts, and the detailed statistics is shown in Tab. 1.

Table 1 Datasets statistics

Datasets	Type	Train	Dev	Test
TCM prescriptions	Sentences	1824	228	228
	Entities	8645	980	900
TCM materia medica	Sentences	2550	318	318
	Entities	7341	880	972

3.2 Baseline

To verify the effectiveness of our approach, it will be compared with the following models. BiLSTM-CRF [5] is one of the baseline models for TCM NER task. BERT-BiLSTM-CRF [4] model is one of the commonly used models for named entity task. Both of the two above models contain the BiLSTM module, which consists of LSTM and backwards LSTM. LSTM [24, 25] is suitable for modelling textual data. LEBERT [26] is integrating lexicon information into the underlying BERT layer and achieved the optimal results for Chinese sequence annotation. Ma et al. [27] proposed to use label information to match entities in text and achieved good results on low-resource Chinese datasets.

3.3 Evaluation Metrics

Our evaluation metrics followed previous work [28], i.e., accuracy (p), recall (R), and $F1$ values [29]. We performed 5 experiments for each implementation and reported their average value as the final experimental result.

3.4 Implementation Details

We used a pre-trained BERTbase (Chinese version) model as an encoder for the character sequences with an initial learning rate of $2e^{-5}$ and a learning rate of $1e^{-4}$ for the other parameters in the model. The detailed hyper parameter settings of the model are shown in Tab. 2.

In the experiment, the maximum number of dictionary tree word matches was set to 3. Also, the matching process discards words consisting of a single character.

Our model was based on the PyTorch architecture and trained using an NVIDIA GeForce RTX 2080Ti (11G) GPU.

Table 2 Experimental parameter settings

Hyperparameters	Numeric
d^c	768
d^{v1}	300
d^k	300
Dropout	0.1
Learning rate (BERT)	$2e^{-5}$
Learning rate (others)	$1e^{-4}$
Batch size	16
Epochs	30
Max length	128

4 EXPERIMENTS

4.1 Comparisons with Baseline Models

Tab. 3 shows the results of our model and the comparison model on the TCM prescription and TCM material medical datasets. It can be observed that our model achieves the optimal F1 value, which indicates that the model is able to effectively utilize not only the lexicon information but also the captured local information to enhance the model performance. The BERT-BiLSTM-CRF and BiLSTM-CRF show that the pre-trained model plays an important role in improving the model performance. The core assumption of Ma et al.'s method is that the label names carry information about the meaning of the labels, and the label names in the TCM dataset are abstract generalizations, thus leading to a lower model recognition accuracy. The LEBERT model outperforms the BERT-BiLSTM-CRF model significantly, proving that the lexicon as well as the word vectors created in this paper are effective. As a result, our model achieved 0.51% and 0.73% improvement on the TCM prescription and TCM material medical datasets improved, respectively.

Table 3 Experimental results

Model	p / %	R / %	$F1$ / %
TCM prescriptions			
BiLSTM-CRF	79.68	78.84	79.25
BERT-BiLSTM-CRF	87.86	90.89	89.35
LEBERT	88.79	92.89	90.62
Ma et al	87.03	90.12	88.55
Ours	89.36	92.97	91.13
TCM materia medica			
BiLSTM-CRF	80.54	82.73	81.62
BERT-BiLSTM-CRF	86.30	89.95	88.06
LEBERT	88.59	90.38	89.48
Ma et al.	86.94	87.89	87.41
Ours	88.76	91.72	90.21

4.2 Ablation Studies

In this section, we carry out detailed experiments on the data sets of traditional Chinese medicine prescriptions and traditional Chinese medicine herbs to prove that each component or design in our model plays a vital role in extracting entity performance methods. As shown in Tab. 4, we use the equation method to describe the removal or replacement of a component in the model. We take BERT-Transformer-CRF as the basis and mark it as M. From Tab. 5, we can draw the following conclusions:

(1) It is reasonable for Transformer Encoder to sense the semantic characteristics of the context. Most previous

work used BiLSTM to capture context information, so we used BiLSTM to replace Transformer Encoder to form a new model, called M1. Observe Tab. 4, variant #2 outperforms #1, so using Transformer Encoder can improve performance.

(2) Lexicon information enhancement is crucial to improve the performance of the model. We use char-word attention to form a variant model based on the basic model, which is recorded as M+CW-ATT. Compared with M+CW-ATT, the F1 value of M decreases by 0.81% and 1.08% on the two data sets, which indicates that dictionary information can improve the performance of the model, and also proves that the character-word attention we proposed is effective.

(3) Local feature extraction is effective. We form a new model M+CW-ATT+TLAE by adding a text local feature extraction (TLFE) module to the M+CW-ATT model. From this, we can see that the F1 value index of #4 model has increased by 0.81% and 0.94% respectively, which indicates the importance of potential word information and local semantic features in TCM NER task.

(4) Complementary Features are crucial. We investigated the effect of complementary fusion gates on model performance. Variant #5 directly integrates context features and local features, and is recorded as M2. Variant #6 only assigns weights to local features, and then performs feature fusion, which is recorded as M3. Variant #7 uses our complementary fusion gate to fuse context features and local features. The results show that Variant #7 using complementary fusion gates achieves the best results, consistent with our intuition about the effect of information redundancy.

Table 4 Results of ablation experiments

Equation	Variant	Datasets	
		TCM prescriptions	TCM materia medica
M	#1	89.51	88.19
M1	#2	89.31	88.06
M+CW-ATT	#3	90.32	89.27
M+CW-ATT+TLFE	#4	91.13	90.21
M2	#5	90.68	89.67
M3	#6	90.80	89.79
M4	#7	91.13	90.21

4.3 Error Analysis

To fully understand the disadvantages and advantages of the model, we performed error analysis. Through experiments, we found that the model was less effective in identifying symptom-like entities in the TCM dataset (Just take the data of TCM prescriptions as an example), and Tab. 5 shows the 3.61% improvement in F1 values for our model over the BERT-BiLSTM-CRF model.

Table 5 Effectiveness of different models for symptom recognition

Model	$F1$ / %
BERT-BiLSTM-CRF	82.74
Ours	86.35

The lack of clear entity boundaries in Chinese text and the relatively complex structure of symptomatic entities with variable entity length lead to poor model recognition results. Although the use of lexicon information and the

acquisition of potential word information can improve the entity recognition effect, how to further capture the boundary information accurately is the goal of the next stage of research.

5 CONCLUSIONS

In this paper, we proposed a TCM NER model with lexicon information and text local feature enhancement to deal with the problem of TCM entity boundaries ambiguity in NER task. Lexicon information is incorporated into the model to enable the model to learn entity boundary information better. And a text local feature extraction module is designed to capture latent lexicon information and local semantic features in sentences. Furthermore, a complementary fusion gate is proposed for adaptively fusing features. Our work provided a good method for practitioners of TCM to automatically extract valuable knowledge from ancient TCM texts. In future work, we will further investigate the TCM entity boundary blurring problem and the fine-grained TCM NER task to obtain knowledge in TCM ancient texts more accurately.

Acknowledgements

This work has been supported by Hebei Province 333 Talent Funding Project "Brain-like Intelligent Knowledge Discovery Technology Research" (Project-number: A201803082).

6 REFERENCES

- [1] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition.
- [2] Xu, H., Liu, H., Jia, Q., Zhan, Y., Zhang, Y., & Xie, Y. (2021). A nested named entity recognition method for traditional Chinese medicine records. *International Conference on Artificial Intelligence and Security*, 488-497.
- [3] Liu, Z., Luo, C., Zheng, Z., Li, Y., Fu, D., Yu, X., & Zhao, J. (2021). TCMNER and PubMed: A Novel Chinese Character-Level-Based Model and a Dataset for TCM Named Entity Recognition. *Journal of Healthcare Engineering*. <https://doi.org/10.1155/2021/3544281>
- [4] Zhang, M., Yang, Z., Liu, C., & Fang, L. (2020). Traditional Chinese medicine knowledge service based on semi-supervised BERT-BiLSTM-CRF model. *2020 International Conference on Service Science (ICSS), Xining, China*, 64-69. <https://doi.org/10.1109/ICSS50103.2020.00018>
- [5] Deng, N., Fu, H., & Chen, X. (2021). Named entity recognition of traditional Chinese medicine patents based on BiLSTM-CRF. *Wireless Communications and Mobile Computing*, 6696205. <https://doi.org/10.1155/2021/6696205>
- [6] Ma, Y., Liu, Y., Zhang, D., Zhang, J., Liu, H., & Xie, Y. (2022). A Multigranularity Text Driven Named Entity Recognition CGAN Model for Traditional Chinese Medicine Literatures. *Computational Intelligence and Neuroscience*, 1495841. <https://doi.org/10.1155/2022/1495841>
- [7] Mahmood, A. T., Kamaruddin, S. S., Naser, R. K., & Nadzir M. M. (2020). A Combination of Lexicon and Machine Learning Approaches for Sentiment Analysis on Facebook. *Journal of System and Management Sciences*, 10(3), 140-150. <https://doi.org/10.33168/JSMS.2020.0310>
- [8] Wang, C., Gao, J., Rao, H., Chen, A., He, J., Jiao, J., ... & Gu, L. (2022). Named entity recognition (NER) for Chinese agricultural diseases and pests based on discourse topic and attention mechanism. *Evolutionary Intelligence*. <https://doi.org/10.1007/s12065-022-00727-w>
- [9] An, Y., Xia, X., Chen, X., Wu, F. X., & Wang, J. (2022). Chinese clinical named entity recognition via multi-head self-attention based BiLSTM-CRF. *Artificial Intelligence in Medicine*, 127, 102282.
- [10] Li, L. & Jiang, Y. (2018). Integrating language model and reading control gate in BiLSTM-CRF for biomedical named entity recognition. *IEEE/ACM transactions on computational biology and bioinformatics*, 17(3), 841-846.
- [11] Li, C., Li, L., & Qi, J. (2018). A self-attentive model with gate mechanism for spoken language understanding. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3824-3833.
- [12] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space.
- [13] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate.
- [14] Li, J., Fei, H., Liu, J., Wu, S., Zhang, M., Teng, C., ... & Li, F. (2022, June). Unified named entity recognition as word-word relation classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(10), 10965-10973.
- [15] Liu, J., Ji, D., Li, J., Xie, D., Teng, C., Zhao, L., & Li, F. (2022). TOE: A Grid-Tagging Discontinuous NER Model Enhanced by Embedding Tag/Word Relations and More Fine-Grained Tags. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 31, 177-187.
- [16] Li, X., Yan, H., Qiu, X., & Huang, X. (2020). FLAT: Chinese NER using flat-lattice transformer.
- [17] Balan, N., & Ila, V. (2022). A Novel Biometric Key Security System with Clustering and Convolutional Neural Network for WSN. *Tehnicki vjesnik-Technical Gazette*, 29(5), 1483-1490. <https://doi.org/10.17559/TV-20211109073558>
- [18] Choi, J. W. & Jeong, E. R. (2022). A multi-output convolutional neural network-based distance and velocity estimation technique. *Journal of Logistics, Informatics and Service Science*, 9(1), 11-25. <https://doi.org/10.33168/LISS.2022.0202>
- [19] Jia, Q., Zhang, D., Yang, S., Xia, C., Shi, Y., Tao, H., ... & Xie, Y. (2021). Traditional Chinese medicine symptom normalization approach leveraging hierarchical semantic information and text matching with attention mechanism. *Journal of Biomedical Informatics*, 116, 103718.
- [20] Lu, J., Yang, J., Batra, D., & Parikh, D. (2016). Hierarchical question-image co-attention for visual question answering. *Advances in neural information processing systems*, 29.
- [21] Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
- [22] Tkachenko, A. (2022), Method for Solving the Linear Multicriteria Optimization Problem in Integers. *Economic Computation and Economic Cybernetics Studies and Research*, 56(1), 159-174. <https://doi.org/10.24818/18423264/56.1.22.10>
- [23] Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory*, 13(2), 260-269.
- [24] Christalin Nelson, S., Tapan Kumar, M., & Prakash G. L. (2022). A Novel Optimized LSTM Networks for Traffic Prediction in VANET. *Journal of System and Management Sciences*, 12(1), 461-479. <https://doi.org/10.33168/JSMS.2022.0130>
- [25] Sujatha, R., Teja, A., Naveen, P., Chatterjee, J. M., Aarthy, S. L., & Hassanien, A. E. (2020). Web Application for Traffic Monitoring and Guidance. *Journal of System and Management Sciences*, 10(4), 1-14. <https://doi.org/10.33168/JSMS.2020.0403>
- [26] Liu, W., Fu, X., Zhang, Y., & Xiao, W. (2021). Lexicon enhanced chinese sequence labeling using bert adapter.

- [27] Ma, J., Ballesteros, M., Doss, S., Anubhai, R., Mallya, S., Al-Onaizan, Y., & Roth, D. (2022). Label Semantics for Few Shot Named Entity Recognition.
- [28] Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging.
- [29] Wang, J., Chen, Y., Qiu, S., & Cui, Q., (2021), Cuckoo Search Optimized Integrated Framework Based on Feature Clustering and Deep Learning for Daily Stock Price Forecasting. *Economic Computation and Economic Cybernetics Studies and Research*, 55(3), 55-70.
<https://doi.org/10.24818/18423264/55.3.21.04>

Contact information:

Yuekun MA

1) School of Computer and Communication Engineering,
University of Science and Technology Beijing, Beijing 100083, China
2) College for Artificial Intelligence,
North China University of Science and Technology, Tangshan 063210, China
3) Hebei Key Laboratory of Industrial Intelligent Perception,
Tangshan 063210, China
E-mail: mayuekun@ncst.edu.cn

He LIU

College for Artificial Intelligence,
North China University of Science and Technology,
Tangshan 063210, China
E-mail: liuh@stu.ncst.edu.cn

Dezheng ZHANG

(Corresponding author)
1) School of Computer and Communication Engineering,
University of Science and Technology Beijing, Beijing 100083, China
2) Beijing Key Laboratory of Knowledge Engineering for Materials Science,
Beijing, China, 100083
E-mail: zdzchina@ustb.edu.cn

Chang GAO

College for Artificial Intelligence,
North China University of Science and Technology,
Tangshan 063210, China
E-mail: 775480184@qq.com

Yujie LIU

College for Artificial Intelligence,
North China University of Science and Technology,
Tangshan 063210, China
E-mail: liuyujie@stu.ncst.edu.cn