Named Entity Recognition Model of Traditional Chinese Medicine Medical Texts based on Contextual Semantic Enhancement and Adversarial Training

Yuekun Ma, Moyan Wen, He Liu

***Abstract*—Traditional Chinese Medicine (TCM) medical texts are the summaries of TCM knowledge and experience, which has practical meaning for the study and analysis of clinical diagnosis. A named entity recognition model is constructed for the field of ancient medical texts in TCM that improves recognition accuracy. The model is based on contextual semantic enhancement and adversarial training. This model designs a contextual semantic enhancement module, which includes a context semantic awareness unit and a complementary enhancement gate unit to capture the context semantic information of sentences and adaptively integrate the original semantic features. On this basis, the robustness of the model is improved by adversarial training. On the data of TCM medical texts, the F1 value of this method reaches 91.95%, which is generally superior to other comparison models.**

***Index Terms*—complementary enhancement gate, enhancement modules, named entity recognition, medical texts**

I. Introduction

As an important part of modern medicine, Traditional Chinese Medicine (TCM) occupies a vital position in daily clinical treatment, such as the prevention and treatment of the new coronavirus. The ancient texts of TCM are the crystallization of the valuable experience of TCM practitioners, contain rich TCM knowledge, and have great reference value for the clinical treatment of TCM. Therefore, extracting value entities from ancient TCM texts is one of the important ways to study and analyze TCM knowledge.

Named Entity Recognition (NER) is responsible for identifying entities with specific significance in unstructured text and categorizing them into categories with expected meanings. including personal names, place names, proper nouns, etc [1]. At present, the NER task in the field of TCM is generally taken as a sequence labeling task, and the model construction architecture is mostly Bi-directional Long-Short Term Memory (BiLSTM) and Conditional Random Field (CRF). Deng et al. used the BiLSTM-CRF to identify entities in TCM patent texts [2]. Zhang' team fused the pre-trained BERT and the BiLSTM-CRF to identify entities in ancient TCM texts [3]. Both Literature [4] and Literature [5] also used the BiLSTM-CRF framework to build entity recognition models. Although the model based on BiLSTM-CRF architecture can effectively extract global context semantic information (SEI), it ignores the correlation dependence between internal information of entities. At the same time, due to the short text sentences of ancient Chinese medicine texts, the entity information relies heavily on contextual SEI. In addition, because labeled data requires rich domain knowledge and good labeling criteria, there is a lack of large-scale and high-quality training data in TCM, which in turn affects the robustness.

Given this, the paper develops a method for NER in TCM medical texts based on contextual semantic enhancement and Adversarial Training (AT).

In this paper, a context feature perception module is designed to learn text semantic features, which consists of two key components: (1) context feature perception unit, which is used to capture the context feature dependency information and global Context Semantic Features (CSFs) inside the entity to guide the refinement of context feature perception; (2) Complementary enhancement gate unit is utilized to adaptively fuse CSFs and original CSFs to achieve context feature perception enhancement. Meanwhile, this paper adds Adversarial Perturbation (AP) to the word embedding vector to improve the model robustness through AT. Lots of experiments on the text dataset of TCM books prove its effectiveness.

II. Related Work

A. Named Entity Recognition

NER is one of the basic tasks of natural language processing (NLP) and an essential part of downstream tasks, like relational extraction [6] and information retrieval [7]. Nowadays, deep learning has become the mainstream approach for Chinese NER models [8]. Huang et al. for the first time used BiLSTM to learn contextual SEI in text and predict labels using CRFs, and the model has become one of NER's infrastructures [9]. Li et al. further acquired contextual SEI by fusing the attention mechanism into the BiLSTM to handle the matter of information loss resulted by long distance [10]. With the introduction of Transformer models [11] such

Manuscript received January 12, 2024; revised July 10, 2024.

This study is supported by "Research on brain-like intelligence knowledge discovery technology", a "333" talent support project of Hebei Province No. A201803082.

Yuekun Ma is a Professor of College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063200, China. (corresponding author e-mail: mayuekun@ncst.edu.cn).

Moyan Wen is a postgraduate student of College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063200, China. (e-mail: wenmy@stu.ncst.edu.cn).

He Liu is a postgraduate student of College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063200, China. (e-mail: liuh@stu.ncst.edu.cn).

as BERT, the performance of Chinese NER models is further improved. For instance, Wu et al. used the BERT as an encoder to generate character vector embeddings and combined them with the BiLSTM-CRF architecture [12]. With the in-depth study of NER, some experts have attempted to incorporate external information into the model to obtain rich contextual SEI. For example, Yang et al. fused the domain dictionary information into the model, making full use of the character in the text and the sequence information of the word, which significantly enhanced the prediction accuracy of the model [13]. Liu et al. designed a word-word adapter to integrate vocabulary into character information, and used the BERT for coding, which achieved the best results in the general domain NER task. Both methods rely on high-quality domain dictionaries, and building dictionaries is extremely time-consuming, so they are not suitable for NER tasks in the field of TCM [14]. In addition, Liu and Lu proposed a cross-domain NER method, which adapts the NER knowledge learned in the high to the low resource-domain to supplement the data SEI in the low-resource domain. However, due to the unique and obscure language structure of ancient TCM texts, it is difficult to transfer knowledge from the general field or other high-resource fields to the field of TCM [15].

B. *Adversarial Training*

AT is one of the crucial ways to enhance the robustness of neural networks. Goodfellow et al. first proposed AT for image classification tasks and improved the robustness by adding small perturbations to the neural network model [16]. AT has shown impressive performance in the field of computer graphics [17]. In recent years, AT is extensively utilized in NLP tasks with good results. Miyato et al. introduced AT into the text classification task, and designed a Fast Gradient Sign Method (FGM) method, which generates perturbations according to the gradient of backpropagation and adds the perturbations to the word embedding vector, thereby enhancing the robustness and improving the overall performance [18]. Wu et al. applied AT to the relation extraction task to lift the model accuracy by adding APs to the word embedding vector [19]. Zhou et al. proposed a two-adversarial migration network based on transfer learning to improve the performance of low-resource NER models [20]. Kitada et al. applied AT to the attention mechanism to improve the predictive performance [21]. Chen et al. designed a novel AT algorithm FreeLB for natural language understanding task. It promoted higher invariance in the embedding space by adding APs to word embedding and minimizing the adversarial risk generated in different regions around the input sample [22]. Therefore, this paper fully excavated the SEI inside the text, and used AT to rise the task accuracy of NER in TCM texts.

III. NER Method of TCM Medical Texts Based on Contextual Semantic Enhancement and AT

To improve the accuracy of NER for TCM texts, a deep learning framework built on BERT and AT was constructed, as shown in Fig. 1. Firstly, through BERT's bidirectional converter structure, dynamic vector encoding at the word level is achieved to capture rich contextual information. Subsequently, AT methods were introduced to enhance the model's robustness to input disturbances by adding noise to the character embedding vector. In the feature perception layer, BiLSTM is used in parallel to capture local details and global contextual features, respectively. During this process, a complementary enhancement gate unit was designed to effectively integrate the original semantic features and context aware features of BERT encoding. Ultimately, CRF is selected as a decoding layer to leverage the dependencies between sequence labels and improve overall annotation performance. The proposed framework minimizes the negative logarithmic likelihood loss function at the sentence level and takes the Viterbi to determine the best label sequence to achieve accurate recognition of entities in TCM texts.

A. *Feature Coding Layer*

a. *BERT encoder*

Firstly, the study investigates the dynamic vector encoding of characters in TCM texts using the bidirectional transformer structure of the BERT model to fully capture contextual information. Compared to traditional static character vectors, the BERT model can significantly enhance the representation ability of character vectors, providing more accurate semantic expression for NER, thereby helping to improve depth of text analysis in TCM ancient books.

Given the $X=\{x_1,\cdots,x_n\}$, n represents the input sentence length. The BERT is used to embed the generated characters into the $E=\{e_{CLS},e_1,\ldots,e_n\}$. The e_{CLS} denotes the SEI representation, as formalized as:

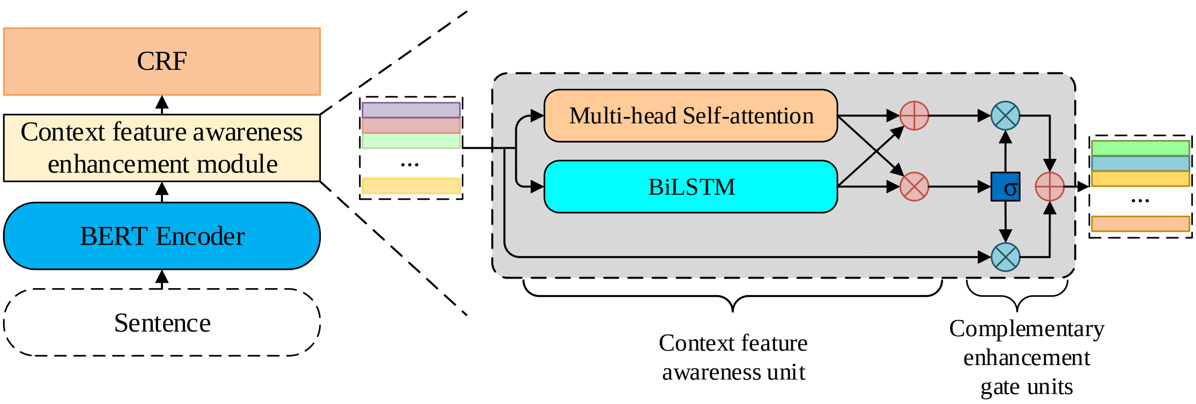

Fig. 1. Model structure.

$$E = BERT(X) \tag{1}$$

b. Adversarial training content

AT is a training method that introduces noise, which enhances the robustness and generalization ability by adding weak interference to the original input to make the model make incorrect predictions. This article adds weak noise to the character embedding vector, rather than the character of the original input, and its model can be abstractly expressed as:

$$\max_{\theta} P(y|x+\Delta x,\theta) \tag{2}$$

In the formula, x is the original input character embedding vector. Δx is the AP. θ represents the model parameters. y is the sample label. $P(\cdot)$ is the probability of predicting the true label after adding interference.

The FGM method is adopted to calculate the AP, and its main idea is to calculate the specific gradient according to the backpropagation to obtain a better AP Δx, and the calculation formula is:

$$\Delta x = \varepsilon g / \|g\|_2 \tag{3}$$

In the formula, ε is the scaling factor. g represents the gradient, and the calculation method of g is:

$$g = \nabla_x L(x,y,\theta) \tag{4}$$

In the formula, $L(x,y,\theta)$ is the loss function.

B. Feature Perception Layer

Subsequently, the feature perception layer was designed. The feature aware layer, as the core of the context feature enhancement module, has two key components: the first is the context feature aware unit, which effectively captures and refines the context dependency relationships of each character in the sentence through BiLSTM; The second is the complementary enhancement gate unit, which can utilize the interaction between the original semantic features and contextual features to adaptively enhance the expression of contextual features, thereby optimizing the NER performance.

a. Context feature awareness unit

(1) Multi-Head Self-Attention Unit

The multi-head self-attention is taken to capture dependencies between contextual semantic features within entities. The single-headed Self-Attention Mechanism (SAM) first calculates the importance between all character feature vectors and generates a new character vector representation by calculating the weighted sum of related features. As the model is trained, the single-head SAM fully exploits the interdependencies within the entities to capture fine-grained contextual semantic features. The multi-head SAM can capture the refined contextual semantic features from different dimensions.

The SAM first maps the input matrix E into three distinct matrices:

$$Q,K,V = EW^Q, EW^K, EW^V \tag{5}$$

In the formula, W^Q, W^K, W^V are the transformation matrices. Then the dot product operation is performed on Q and K, and normalized using the SoftMax function, and finally the attention matrix is obtained:

$$Atten(Q,K,V) = SoftMax\left(Q^T K\right)V \tag{6}$$

The SAM is simplified here to improve the model efficiency. The long SAM is a splicing of multiple SAMs, and its formula is:

$$MultiHead\ (Q,K,V) = Concact(\ head\ _1,\ldots,\ head\ _n) \tag{7}$$

$$head\ _i = Atten(Q_i, K_i, V_i) \tag{8}$$

Finally, the dimension of the output feature matrix W_P^1 is converted into the same dimension as the input matrix by transforming the matrix. The output of the bull's SAM, H_{MA}, can be shown as:

$$H_{MA} = MultiHead(Q,K,V)W_P^1 \tag{9}$$

It is worth noting that the global semantic feature vector produced by the BERT is represented by e_{CLS} and used as the input of the multi-head SAM, which has the advantage of using the global SEI to lift the character information at each position.

(2) BiLSTM Network

Each entity in the sentence is implicitly included in all character features. Therefore, the global context SEI for an entity helps to NER tasks. The BiLSTM model can better capture bidirectional SEI and is adopted in NER tasks. This paper follows previous work to extract global context SEI using the BiLSTM model. The output H_{CFE} of BiLSTM can be expressed as:

$$H_{CFE} = BiLSTM(E)W_P^2 \tag{10}$$

In the formula, W_P^2 is the transformation matrix, which has the same function as W_P^1.

(3) Feature integration

Feature H_{MA} and Feature H_{CFE} contain different information. The two features are integrated in the following ways:

$$H = H_{MA} \oplus H_{CFE} \tag{11}$$

In the formula, $\oplus$ represents the sum calculation of the corresponding elements, and H represents the context semantic characteristics after integration. In addition, in this paper, the point product operation of H_{MA} and H_{CFE} is performed to generate the weight information W_b, which is used to complement the enhancement gate element, and the calculation formula is as follows:

$$W_b = H_{MA} \otimes H_{CFE} \tag{12}$$

In the formula, $\otimes$ represents the calculation of the dot product of the corresponding element.

b. Complementary enhancement gate unit

The original SEI generated by BERT also contains rich contextual SEI, so this paper designs a complementary enhancement gate to make full use of the original SEI, which can adaptively integrate the original semantic feature E into the contextual semantic feature H.

The input of the complementary enhancement gate unit has three parts: the original semantic feature E, the CSF H, and the weight information W_b. The output of the complementary enhancement gate unit is:

$$\tilde{E} = H \otimes \sigma(W_b) + E \otimes (1 - \sigma(W_b)) \tag{13}$$

In the formula, $\otimes$ represents the dot product calculation of the corresponding element, and σ is the sigmoid function.

The previous TCM NER model has not used the original SEI. The complementary enhancement gate unit proposed in this paper aims to supplement the context SEI by using the original SEI to further enhance the expression ability of the CSF representation. This paper considers assigning weights to two feature vectors at the same time, rather than assigning weights to the original feature vectors separately, which has the advantage of achieving an adaptive balance between the original SEI and the context SEI and avoiding information redundancy.

C. CRF Decoding Layer

Finally, to capture and model the dependencies of label sequences, the study uses CRF as the decoding layer to ensure the coherence and consistency of predicted labels in the sequence. The CRF layer can optimize the joint probability of the entire label sequence by considering the transition probability between adjacent labels, thereby enhancing the performance in sequence annotation tasks. Firstly, the output $\tilde{E}$ dimensionality reduction of the feature perception layer is used by the linear change layer, and the label score O is calculated:

$$O = W\tilde{E} + b \tag{14}$$

In the formula, W and b are the linear transformation layer's parameter and bias parameter. For the $Y = \{y_1, \ldots, y_n\}$, this paper defines its probability as:

$$p(y|x) = \frac{\exp\left(\sum_i \left(O_{i,y_i} + T_{y_{i-1},y_i}\right)\right)}{\sum_y \exp\left(\sum_i \left(O_{i,\tilde{y}_i} + T_{\tilde{y}_{i-1}}, \tilde{y}_i\right)\right)} \tag{15}$$

In the formula, T is the transition matrix. $\tilde{y}$ represents the candidate tag sequence.

D. Training Target

To train the research model by minimizing the inactive log likelihood loss at the sentence-level. The $\{x_j, y_j\}, 1 \le j \le N$ calculation loss for the given training data is:

$$L = -\sum_j \log\left(p\left(y_j \middle| x_j\right)\right) \tag{16}$$

At last, the optimal label sequence is calculated using Viterby's algorithm [23].

IV. Experiment

A. Data Set

This paper used the "Yellow Emperor's Inner Classic" and "Treatise on Cold Damage Disorders". The "Yellow Emperor's Inner Classic" is known as the "ancestor of medicine", and there were a large number of physical categories such as TCM symptoms, disease names, and TCM signs, and they are elaborately elaborated. "Treatise on Cold Damage Disorders" has established a unique knowledge system of TCM, and this paper used it as a comparison to verify the effect of the model on NER in large-scale corpus. According to the knowledge architecture of TCM, the entity categories of TCM were divided into TCM symptoms, TCM disease names, TCM symptoms, TCM symptoms (including symptoms and pulses), prescriptions, dosages and other entity concepts. Under the guidance of TCM experts, 4852 sentences were manually labeled, which were divided into training sets, verification sets and test sets according to the ratio of 8:1:1. The TCM text dataset contained a total of 6 types of entities, namely "prescription", "Chinese medicine", "symptoms", "pulse", "tongue", and "dose". The detailed statistics are listed in Table I.

B. Model Evaluation and Model Setup

To fully assess the model performance, this paper used the following 3 evaluation metrics: Precision (P), Recall (R), and F1 values. In general, the F1 value is used to evaluate the overall performance, and its formula is:

$$F1 = \frac{2 * P^* R}{P + R} \tag{17}$$

The pre-trained model used was BERT-based (version Chinese), which had a learning rate of 1e-5, while that of CRF in the model was 1e-3, and the remaining other parameters had a learning rate of 1e-4. The model as a whole used the Adam optimizer to update parameters. The experiments use the NIVIDIA GeForce RTX 2080Ti (11G) GPU.

C. Experimental Results and Analysis

To test the performance of the NER of TCM texts, the results were compared with BiLSTM-CRF, BERT-BiLSTM-CRF, RoBERTa-BiLSTM-CRF, RoBERTa dynamic fusion model [24] and AT-RBC model [25] under the same experimental environment, as shown in Table II.

From the observation in Table II, the NER based on CSFs obtained the optimal P value and F1 value on the text dataset of ancient TCM books compared with other comparison models. Experimental results showed that the contextual semantic enhancement module designed in the proposed model could fully mine the contextual SEI in the sentence, and could also supplement the contextual SEI with the original SEI, which effectively improved the overall performance. In addition, AT was used to improve the recognition robustness in the scenario of low data resources.

To further test the performance in lower resource conditions, the experiment was tested on the 10% and 50% scale training sets, respectively, as shown in Table III.

From the observation in Table III, the NER model could still achieve optimal performance in low-resource scenarios, especially the accuracy rate was 1.77% higher than that of the optimal comparison model on the 10% training set, indicating that the model had better practical value in the task of naming entity recognition in ancient TCM texts.

D. Ablation Experiments

To fully understand the impact of each component on model performance, ablation experiments were performed for analysis.

Table I
Dataset statistics

Dataset	Quantity of sentences	Amount of entities
Training	3600	5712
Validation	450	1224
Test	450	1145

Table II
Results of the proposed model and the comparison model

Model	P(%)	R(%)	F1(%)
BiLSTM-CRF	86.13	87.00	86.56
BERT-BiLSTM-CRF	89.30	92.08	90.67
RoBERTa-BiLSTM-CRF	89.40	92.45	90.90
RoBERTa dynamic fusion	89.73	93.25	91.45
AT-RBC	89.25	93.03	91.24
Research model	91.03	92.88	91.95

Table III
Comparison of different models in the low-resource condition

Model	P(%)	R(%)	F1(%)
10% training set			
BiLSTM-CRF	74.37	76.69	75.51
BERT-BiLSTM-CRF	80.20	86.20	83.09
RoBERTa-BiLSTM-CRF	83.67	84.51	84.09
RoBERTa dynamic fusion	82.88	87.87	85.30
AT-RBC	83.21	86.71	84.92
Research model	85.44	86.93	86.18
50% training set			
BiLSTM-CRF	84.77	86.49	85.62
BERT-BiLSTM-CRF	88.06	90.56	89.29
RoBERTa-BiLSTM-CRF	88.54	91.43	89.96
RoBERTa dynamic fusion	89.22	91.36	90.28
AT-RBC	88.32	91.72	89.99
Research model	90.58	91.50	91.04

"Primitive semantics only": The model only took the raw feature vectors produced by the BERT for entity prediction. In other words, the model was a BERT-CRF.

"Context-only semantics": The model did not supplement the CSFs with the original features generated by the BERT model.

"No complementary enhancement gate used": The model only added the original features generated by the BERT to the contextual semantic features.

"AT not used": The model did not use AT.

The comparison results between the presented model and other ablation models are displayed in Table IV.

Table IV
Results of ablation studies with different components

Model	F1(%)
"Primitive semantics only"	90.51
"Context-only semantics"	90.85
"No complementary enhancement gate used"	91.09
"AT not used"	91.75
Research model	91.95

Table IV shows that each component contributes to semantic understanding and entity recognition tasks. When the model is limited to using only original semantics, its F1 score is 90.51%, which verifies the basic ability of the BERT in capturing semantic features. When the model only relies on "pure contextual semantics," its performance slightly improves to 90.85%, indicating the importance of contextual information for understanding texts. However, a more significant performance improvement occurred in models that did not use complementary enhancement gates, with an F1 score of 91.09%, confirming the effectiveness of combining original features with contextual features. In addition, the model that did not use AT further improved its entity recognition performance to 91.75%, reflecting the role of AT in enhancing the model's ability to resist disturbances. In the end, the complete research model highlighted the performance optimization under the synergistic effect of various components with a max-F1 score of 91.95%. These results collectively demonstrate the importance of each part of the model in improving overall recognition performance, as well as their synergistic effects in building efficient entity recognition systems. The performance in the NER task of ancient TCM texts is shown in Fig. 2.

In the scenario of collating ancient medical books, the model proposed in the study achieved an F1 value of 93.03%, showing its efficiency in processing classic documents such as the "Huangdi Neijing". In terms of TCM prescription analysis, the model also performed well, with an F1 value of 92.72%, which showed that the model could effectively extract medicinal materials and dosage information from prescription descriptions. The results of the clinical record mining scenario showed that the model had high P and R in extracting key information such as diseases, symptoms, and treatment methods, with an F1 value of 96.98%. In addition, in the scenarios of knowledge extraction from Chinese medicine literature and correlation research of Chinese medicine symptoms, the model also showed good performance, with F1 values of 94.78% and 95.55%, respectively.

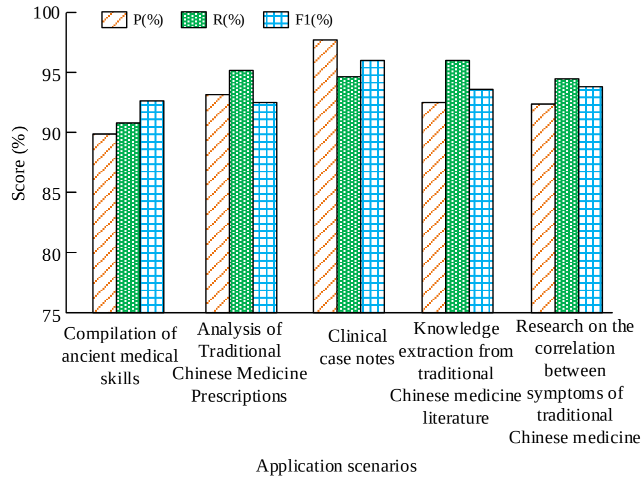

Fig. 2. Comparison of NER performance in different application scenarios.

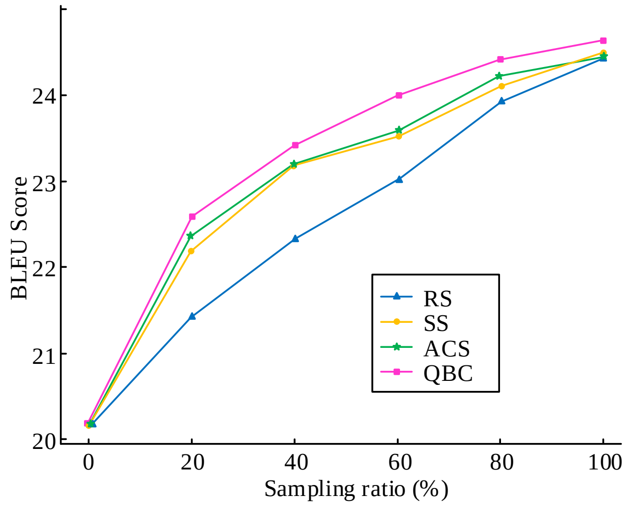

Fig. 3. Active learning algorithm experimental results chart.

To evaluate the effectiveness of active learning strategies in TCM text NER tasks, this study compared three active learning sampling strategies: sentence similarity (SS), attention-based word (ACS), and voting committee (QBC), and compared them with random sampling (RS). Using a 1-million-sample training set, the study aimed to optimize model performance and efficiency through selective incremental training with SGD optimization and a Chinese-to-English language focus. The results are exhibited in Fig. 3.

In Fig. 3, active learning algorithms could significantly improve learning efficiency in the early stages of model training, especially on relatively limited data sets. Active learning allowed the model to quickly learn and improve performance in the early stages of experiments by prioritizing the samples that are most helpful for improving the model for incremental training. However, as the experiment progressed, the number of samples that can greatly improve the performance, which led to the gradual reduction of the benefits obtained by the model through active learning strategies. At the same time, the RS strategy did not perform sample screening, so the learning efficiency of the model remained relatively stable throughout the entire experiment without significant fluctuations. This result indicates that active learning strategies are of great significance for constructing efficient and accurate NER models for TCM texts, which can achieve fast and effective entity recognition under limited resources.

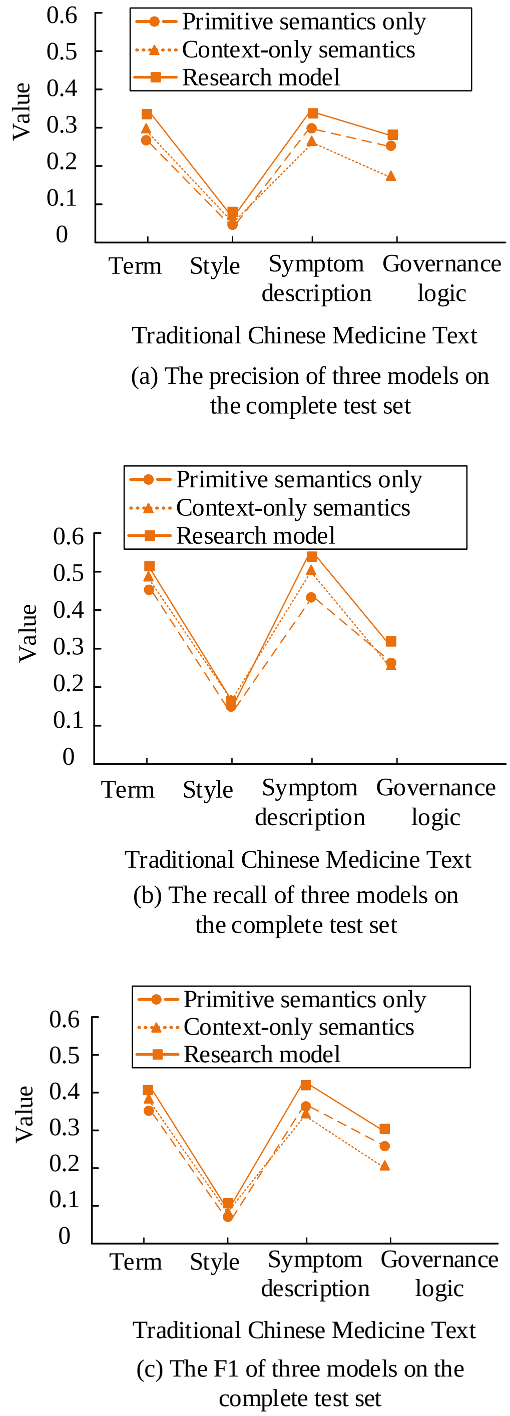

Fig. 4. Comparison results of accuracy, recall rate and F1 value of the three methods on the test set of "Shanghan Lun".

Fig. 4 shows the comparison results of accuracy, recall, and F1 values of three methods on the test set of Shanghan Lun. Among them, the accuracy of the research model for Governance logic is 8.97% and 18.41% higher than that of Primitive semantics only and Context only semantics, respectively. The recall rates are 5.55% and 5.77% higher, respectively; The Fl values were 7.65% and 12.78% higher, respectively. Based on the above experimental analysis, "Primitive semantics only" has a good recognition effect on TCM vocabulary, while "Context only semantics" has advantages in recognizing TCM sentences. Combining these two methods can effectively improve the recognition rate of TCM text content.

V. Conclusion

Aiming at the fact that the TCM NER model does not consider the internal dependence of the entity, a NER method for TCM records grounded on contextual semantic enhancement and AT was constructed. This method designs a contextual semantic enhancement module to inherit the relationship between internal features of entities and global CSFs. In addition, the module obtains complete contextual SEI by combining the original semantic features with the contextual semantic features. In addition, the model AT is adopted to enhance robustness and generalization. The results demonstrated that the proposed method can well identify entities in of ancient TCM texts, and the model's performance is better than that of the comparison model. Future work will further explore the task of Named Entity in fine-grained ancient TCM texts.

References

[1] Md. Asfi-Ar-Raihan Asif, Mirza Muntasir Nishat, Fahim Faisal, Rezuanur Rahman Dip, Mahmudul Hasan Udoy, Md. Fahim Shikder, and Ragib Ahsan, "Performance Evaluation and Comparative Analysis of Different Machine Learning Algorithms in Predicting Cardiovascular Disease," Engineering Letters, vol. 29, no.2, pp731-741, 2021.

[2] N. Deng, H. Fu, and X. Chen, "Named entity recognition of traditional Chinese medicine patents based on BiLSTM-CRF," *Wireless Communications and Mobile Computing*, vol. 6696205, pp. 1-12, 2021.

[3] Carmine Maria Pappalardo, Antonio Lettieri, and Domenico Guida, "Identification of a Dynamical Model of the Latching Mechanism of an Aircraft Hatch Door using the Numerical Algorithms for Subspace State-Space System Identification," IAENG International Journal of Applied Mathematics, vol. 51, no.2, pp346-359, 2021.

[4] H. Miyajima, N. Shigei, H Miyajima and N. Shiratori, "Securely Distributed Computation with Divided Data for Particle Swarm Optimization," Lecture Notes in Engineering and Computer Science: Proceedings of The International Multi Conference of Engineers and Computer Scientists, pp, 20-22, 2021.

[5] Y. Ma, H. Liu, D. Zhang and Y. Liu, "A Named Entity Recognition Method Enhanced with Lexicon Information and Text Local Feature," *Tehnički vjesnik*, vol. 30, no. 3, pp. 899-906, 2023.

[6] Z. Zhong and D. Chen, "A Frustratingly Easy Approach for Entity and Relation Extraction," In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 50-61, 2021.

[7] J. Lin, "A proposed conceptual framework for a representational approach to information retrieval," ACM SIGIR Forum. New York, USA., vol. 55, no. 2, pp. 1-29, 2022.

[8] J. Li, A. Sun, J. Han and C. Li, "A survey on deep learning for named entity recognition," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50-70, 2020.

[9] Z. Huang, W. Xu and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," arXiv:1508.01991, 2015.

[10] L. Li, J. Zhao, L. Hou, Y. Zhai, J. Shi and F. Cui, "An attention-based deep learning model for clinical named entity recognition of Chinese electronic medical records," *BMC Medical Informatics and Decision Making*, vol. 19, no. 5, pp. 1-11, 2019.

[11] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[12] J. Wu, D. Shao, J. Guo, Y. Cheng and G. Huang, "Character-Based Deep Learning Approaches for Clinical Named Entity Recognition: A Comparative Study Using Chinese EHR Texts," In Proceedings of international conference on Smart Health, Springer, Cham, pp. 311-322, 2019.

[13] J. Yang, H. Wang, Y. Tang and F. Yang, "Incorporating Lexicon and Character Glyph and Morphological Features into BiLSTM-CRF for Chinese Medical NER," In Proceedings of 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guanzhou, pp. 12-17, 2021.

[14] W. Liu, X. Fu, Y. Zhang and W. Xiao, "Lexicon Enhanced Chinese Sequence Labeling Using BERT Adapter," In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, pp. 5847-5858, 2021.

[15] B. Y. Lin and W. Lu, "Neural Adaptation Layers for Cross-domain Named Entity Recognition," In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, pp. 2012-2022, 2018.

[16] I. J. Goodfellow, J. Shlens and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv:1412.6572, 2014.

[17] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson, Z. Celik and A. Swami, "The Limitations of Deep Learning in Adversarial Setting," Proceedings of 2016 IEEE European Symposium on Security and Privacy (EuroS&P), Saarbrucken, pp. 372-387, 2016.

[18] T. Miyato, A. M. Dai and I. Goodfellow, "Adversarial training methods for semi-supervised text classification," arXiv:1605.07725, 2016.

[19] Y. Wu, D. Bamman and S. Russell, "Adversarial Training for Relation Extraction," Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, pp. 1778-1783, 2017.

[20] J. T. Zhou, H. Zhang, D. Jin, H. Zhu, M. Fang, R. Goh and K. Kwok, "Dual Adversarial Neural Transfer for Low-Resource Named Entity Recognition," Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, pp. 3461-3471, 2019.

[21] S. Kitada and H. Iyatomi, "Attention meets perturbations: Robust and interpretable attention with adversarial training," IEEE Access, vol. 9, pp. 92974-92985, 2021.

[22] C. Zhu, Y. Cheng, Z. Gan, S. Sun, T. Goldstein and J. Liu, "Freelb: Enhanced adversarial training for natural language understanding," arXiv:1909.11764, 2019.

[23] A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260-269, 1967.

[24] Y. Zhang and Y. Wang, B. Li, "Identifying Named Entities of Chinese Electronic Medical Records Based on RoBERTA wwm Dynamic Fusion Model," Data Analysis and Knowledge Discovery, vol. 6, no. Z1， pp. 242-250, 2022.

[25] R. Guo and H. Zhang, "Chinese medical named entity recognition based on RoBERTa and adversarial training," *Journal of East China University of Science and Technology*, vol. 49, no. 1, pp. 144-152, 2023.