LEVERAGING TRANSFER LEARNING AND LABEL OPTIMIZATION FOR ENHANCED TRADITIONAL CHINESE MEDICINE NER PERFORMANCE

HUANG ZIKUN
SAIDAH SAAD

ABSTRACT

Named Entity Recognition (NER) plays a vital role in various domains, including medical and financial fields, by identifying text fragments belonging to predefined categories from unstructured text. Over time, NER algorithms have evolved from dictionary-based approaches to machine learning and deep learning techniques. Transfer learning, a novel deep learning method, has recently demonstrated impressive results in NER tasks. However, transfer learning models still encounter challenges such as limited entity labels and the impact of noisy datasets. Therefore, this project aims to optimize applying deep learning models for NER and achieve enhanced results. Initially, the BERT+CRF model was applied to the WanChuang dataset, resulting in an F1-Measure of 89.1%, establishing the feasibility of using transfer learning models for NER on Chinese medical data. This model served as a baseline for comparison in our project. To address label-related issues in the baseline model, we proposed a scheme to improve the learning rate of the CRF layer, resulting in an increased F1-Measure of 91.0%. Additionally, to mitigate the impact of noisy training data, we introduced a 10-fold retraining scheme to optimize the training set. By retraining the model using the optimized training set, we achieved an optimal F1-Measure of 92.7%. The experiments demonstrated that the transfer learning model enhances NER entity extraction capabilities while the optimized CRF layer effectively captures the internal relationships of entity tags, thus improving overall performance.

Keyword: Named Entity Recognition, Traditional Chinese medicine, transfer learning, BERT, CRF

INTRODUCTION

Named entity recognition (NER) is of great significance in Natural language processing. NER mainly processes structured and unstructured data, dividing these named entities into predefined classes. Traditional Chinese medicine (TCM) literature NER is an important work in knowledge extraction in the field of traditional Chinese medicine, which refers to extracting concept instances from large-scale unstructured traditional Chinese medicine literature and determining concept types. In the field of TCM, there are a large number of ancient books and medical records, which contain a large number of clinical terms of TCM. These terms contain rich and high-value information. Information includes medicines, symptoms, and diseases, which is conducive to the construction of TCM expert system, TCM knowledge map, and TCM question answering system(Yu et al. 2017).

Before the advent of deep learning methods, traditional NER methods for both general domain texts and TCM texts were deficient. NER methods based on traditional dictionary methods are difficult to scale and optimize. Because the domain characteristics of TCM texts are not considered, the generalization ability of the NER method for TCM texts in general domain texts is weak, and the recognition results are not satisfactory. The NER algorithm based on machine learning requires manual definition of templates and adjustment of parameters. Compared with other methods, the NER model based on deep learning has achieved better results in the field of TCM texts.

Learning models based on deep migration have achieved great success in the field of Natural language processing(Luo et al. 2020). These models can use text data to embed words in context and achieve better accuracy in multilingual understanding tasks. Among them, BERT uses masking language modeling to pre train the bidirectional encoder in a large-scale universal domain corpus for sentence prediction, and the expression effect is the most significant. This model provides a pre trained model for Chinese, which can be better transferred to the field of TCM. Therefore, this study believes that applying BERT to NER tasks in the field of TCM can achieve better results, and its performance is evaluated by training on dataset.

This study suggests that using transfer learning algorithms for NER tasks is a necessary and more optimized option. Transfer learning uses existing knowledge to learn new knowledge The experience of transfer learning is to identify similarities between existing knowledge and new knowledge, and to leverage these similarities to facilitate the transfer of knowledge In the field of TCM, corporate and annotated data are scattered Meanwhile, due to knowledge limitations in domain specific terms in medical field, data annotation of cushions from noise In order to solve the large amount of training data required for in depth learning training, Transfer learning is used to pre learn the knowledge in the public text field, and then transferred to the field of traditional Chinese medicine Researchers have used this method to perform NER tasks based on transfer learning models such as BERT and achieved Excel results(Zhang et al. 2022).

This paper consists of five sections. The first section discusses the background of this study, including the problems in TCM NER. The second section summarizes historical research and NER methods. The third section elaborates on the methods used in the study. The fourth section introduces the results of the work and discussion. Finally, the fifth section summarizes the research results and suggests future work.

LITERATURE REVIEW

A. CHINESE TRADITIONAL MEDICINE (TCM)

Traditional Chinese Medicine (TCM) generally refers to a kind of traditional medicine created by the Han nationality in China. The study of TCM encompasses human physiology, pathology, diagnosis, treatment, prevention, and management of diseases. TCM originated in primitive societies and developed its theoretical framework during the Spring and Autumn and Warring States periods. In addition, TCM has significantly influenced other countries with Chinese character civilizations, such as Japanese Chinese medicine(Cheng 2014), Korean medicine(Ju-Ah et al. 2017), Vietnamese medicine(Bui 2019) and so on. TCM has a long history, has a complete theoretical system and unique treatment, has accumulated a lot of clinical experience. There is a particularly rich reserve in TCM prescriptions and literature on treating diseases.

With the development of information technology construction, post structured research on TCM related information has become a mainstream trend. Record the ingredients, dosage, and corresponding treatment symptoms of traditional Chinese medicine prescriptions. Using this information, the diagnosis treatment experience and compatibility plans of TCM can be summarize and providing reference for future learning and use. It is one of the important tasks in the current medical field to extract medical entity information from texts using electronic medical records and Chinese herbal medicine text information as text corpora and natural language processing related technologies(Wang et al. 2020). NER work is an important and crucial foundational step that can serve as a tool for traditional Chinese medicine artificial intelligence to assist clinical decision-making, with enormous theoretical and applied research value. In the field of modern TCM, we mainly face some problems. Although there are certain norms in the expression of medical terminology, as a natural language, it is still a free text expression. Different doctors often use different expressions of traditional Chinese medicine terminology when expressing the same meaning(Lei et al. 2014). Due to the long history of TCM development, the descriptions and terminology of the same drug may be different in different historical documents, and drugs with similar names may also be quite different(Sarkar et al. 2023). This makes the dissemination and extraction of information difficult. It is necessary to propose a scheme that can uniformly extract the diseases in the literature.

In summary, it is meaningful to propose a work that does not rely on manual recognition and can complete NER tasks in TCM.

B. NER FOR TRADITIONAL CHINESE MEDICINE

NER in the field of TCM refers to the application research of NER in the field of traditional Chinese medicine, it refers to the understanding of multiple entities with specific significance in the text of TCM, such as "cold" classified as illness, "cough" classified as symptom, and "ginseng" classified as drug. TCM entity is the basic element of TCM text, and the exploration of its internal relationship is the essential work of TCM research. The traditional methods of TCM research are mainly manual extraction of unstructured texts such as TCM literature, TCM clinical diagnosis records, etc. In recent years, the rapid advancement of natural language processing (NLP) technology has led to an increasing integration of NER and TCM. Employing NER technology, as opposed to traditional manual methods, to extract traditional Chinese medicine ingredients can significantly reduce the time and manpower required, thereby enhancing the efficiency of Chinese medicine ingredient research.

The NER task in the field of traditional Chinese medicine can be seen as a sub domain NER task. The traditional Chinese medicine NER task is mainly based on a Chinese text corpus. Public traditional Chinese medicine societies, such as CMeEE. However, because most medical NER tasks need to focus on more subdivided fields in their own medicine, such as diabetes, more medical NER literatures choose to use their own data sets.

C. NER METHOD FOR CHINESE LANGUAGE

we divided the existing NER methods into rule-based methods, statistical machine learning methods, and deep learning methods.

i. Rules Based Method

previous studies mainly used rules and dictionaries to identify named entities, which were constructed manually. Select keywords as features and use pattern matching to extract text corresponding entities. This approach is limited by the need for manual rules to differentiate entity types and its lack of portability. The effectiveness of this method depends on the capacity of the dictionary and pattern rules(Krstev et al. 2014). Furthermore, constructing a large dictionary can be challenging, and updating it requires time and effort. Although this method appears straightforward, it has limitations. Rule-based systems mainly rely on manually crafted semantic and grammatical rules to identify entities, which is limited by dictionary capacity and results in high accuracy but low recall(Chen et al. 2020). Analyzing the lexical features and collocation habits of named entities, construct artificial recognition rules for named entities. In this process, there is a need for continuous improvement and completion of rules. But each rule should be written before actual use to solve the problem of multiple named entities using context.

The advantage of using a dictionary for entity recognition is that it produces almost no errors, as all vocabulary is already included in the dictionary. On the other hand, he has little ability to make predictions based on dictionaries, because the algorithm cannot recognize content beyond the dictionary. Grammatical rule-based algorithms can recognize parts beyond the dictionary, but in TCM, the composition of prescriptions is not entirely consistent with the English or Chinese grammar used by humans.

ii. Machine Learning Based Method

NER tasks based on machine learning are divided into supervised learning and unsupervised learning. In supervised learning, NER can be transformed into multi-class and sequential tagging tasks. In a tagged data sample, each sample can be represented by a carefully constructed feature. On this basis, machine learning method is used to model the labeled data, and the trained model is used to recognize the unknown data quickly(Bin et al. 2016). Currently, various machine learning methods, such as Hidden Markov Models (HMM), Support Vector Machines (SVM), and Conditional Random Fields (CRF), are commonly employed in supervised NER tasks. Unsupervised methods, primarily based on clustering techniques or similarity judgments between entities and seed terms, are also used. Entity recognition is achieved through statistical analysis using lexical features on large-scale unlabeled corpora. Different text clusters are obtained according to text similarity, representing different entity groups. Commonly used features or auxiliary information include lexical resources, Term Frequency-Inverse Document Frequency (TF-IDF), shallow semantic information (blocked NP-chunking). The method of applying machine learning to solve NER tasks is widely used because of the support of mathematical logic and strong interpretability.

iii. Deep Learning Based Method

With the development of computer hardware and word embedding technology, neural networks have been able to solve many natural language processing problems. This method can also be applied to NER: the discrete word expression is mapped to low-dimensional space, the dense word is obtained, then the word order is entered into recursive neural network, the feature is extracted by neural network and the maximum value makes predicted. Compared to linear HMM and CRF, deep learning algorithm can extract features from original data. The deep learning solution for NER tasks is mainly divided into three steps: distributed text representation, context encoding architecture, tag decoder.

D. OPTIMIZATION

The optimization scheme is mainly divided into two aspects: data optimization and model optimization.

Data optimization in the context of NER refers to improving the quality and effectiveness of training data used to train NER models. It involves various techniques aimed at enhancing the data to enhance the model's performance and generalization capabilities. Data optimization techniques involve such as data cleaning, label consistency, data augmentation, accurate-consistent annotation and so on. Whether it is supervised learning or unsupervised learning, data is always the most important driving force. More types of data can bring better stability and predictability to good models of unknown data. For the model, the data encountered is more likely to be recognized than the previously unseen data. But adding data is not blind.

Model optimization is mainly aimed at the modification of the model itself or the adjustment of hyperparameters. Training a neural network model involves the utilization of optimization algorithms to solve the parameter optimization problem by minimizing the cost function. Training a neural network model may require several hundred or thousands of machines to train simultaneously for several months. Using optimization algorithm can save training time to accelerate the model convergence. Common optimization algorithms include gradient descent algorithm, RMS prop algorithm(Bekoulis et al. 2018) and Adam algorithm(Roth and Hongxia 2016). Using random gradient descent can help us to try to converge to the global optimal value and get rid of local minimum and saddle point. But this is only a theoretical case, no one algorithm can guarantee in such a large number of parameters to find the optimal solution.

RESEARCH MODEL AND RESEARCH QUESTIONS

The NER task is still an open research field in the field of Chinese NLP, because it brings many challenges, including the migration and scalability of NER algorithms, and the problem of noisy data annotation in domain specific term in medical field.

This project employs the BERT+CRF transfer learning model as a baseline and applies it to the NER task. The learning rate of the CRF layers is independently adjusted to achieve improved results within a limited number of training epochs. Additionally, to address the data noise problem, a 10-fold training method is proposed to optimize the training set and enhance the model's recognition effectiveness.

The research scope of this project is the application and improvement of a transfer learning model based on Chinese traditional Chinese medicine dataset. The focus of this study is to use the deep learning model BERT+CRF for NER tasks based on the WanChuang dataset. As far as the improvement content of this project is concerned, these include increasing the Learning rate of CRF layer and 10-flod retraining mode. The following are the questions that will be investigated in this project:

i. How to improve the learning ability of the CRF layer in the BERT+CRF model structure to improve the restriction ability of the entity relationship?
ii. How to reduce the impact of dataset noise without expert knowledge?

The main purpose of this study is to practice the feasibility of the BERT+CRF model on the NER task and improve it. The objectives of the research are as follows:

i. To the BERT+CRF model to perform NER tasks on the Chinese traditional Chinese medicine dataset to verify its feasibility and serve as a baseline model.
ii. To proposed optimizations for the learning rate of the CRF layer in the baseline model and compare the experimental results after retraining.
iii. To proposed the 10-fold optimization scheme to optimize the noise of the training set and compare it with the experimental results after retraining.

METHODSLOGY

The flowchart shown in Figure 1 is the process of the method used in this study. The NER task of this study is based on the BERT+CRF model and is divided into 7 steps for model training. This includes dataset preprocessing, improve CRF learning rate, 10-fold label optimization, content embedding, training of BERT model and CRF model, as well as final prediction output and evaluation. This study modifies the parameters based on the prediction obtained from the baseline model BERT+CRF model structure and retrains them to obtain better results.

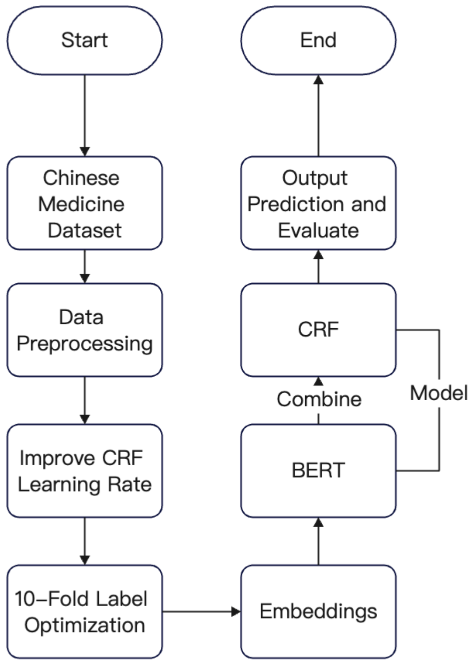

FIGURE 1. Overview of Methodology

The first step includes data collection and preprocessing. Based on literature references and the collection of publicly available datasets, and based on the focus of this project, we have chosen publicly available dataset Wanchuang(Tianchi 2020) as our dataset for our work. A publicly available dataset is more conducive to comparing and evaluating our work with other studies. In this step, we will preprocess the dataset based on its content, including data cleaning, segmentation, and evaluation.

The second stage is to evaluate based on the data processed in the previous stage. Based on the model super parameters and label errors and omissions in the dataset, the project first improved the learning rate of the CRF layer, and then used a 10-fold method to supplement and delete the labels in the training set to achieve better NER task performance.

The third step is to embed the dataset through a converter, including word embedding, statement embedding, and positional embedding. The purpose of this step is to transform the natural language dataset into the input part of the model through embedding for BERT model training.

The fourth step involves training the model, which includes fine-tuning the pre-trained BERT model and training the CRF model. Model training was conducted using the Colab Notebook, an online work platform provided by Google, offering

GPU computing power. The "bert-base-chinese" branch was chosen for this Chinese dataset, leveraging the pre-trained BERT model to achieve improved results. The BERT model captures label relationships in the dataset, while the CRF layer ensures predicted values adhere to logical constraints.

Following model training, a validation set is utilized to assess the model's effectiveness and compare it with the baseline BERT+CRF model. This project primarily examines the influence of different annotated datasets on experimental outcomes. Finally, all experimental records and results are summarized and compared, culminating in a comprehensive research summary.

RESULTS AND DISCUSSION

All model training was conducted using Google Collab laptops, leveraging the advantages of powerful GPU acceleration. The training process for 30 epochs is only completed within 1 hour. Initially, the baseline model was trained for 30 epochs with a Learning rate of 1e-6. F1-Measure were evaluated in each round of training, including on the current and test sets, as performance indicators. It can be seen from Figure 2 that the F1-Measure on the test set has reached stability over 20 epochs. The loss curve in Figure 3 indicates that the loss value has not continued to decrease. The experimental results of the baseline model indicate that training for more than 20 epochs is unnecessary. The baseline model achieves the highest F1-Measure of 89.1% at training epoch 15 out of 30 training epochs and remains stable thereafter.

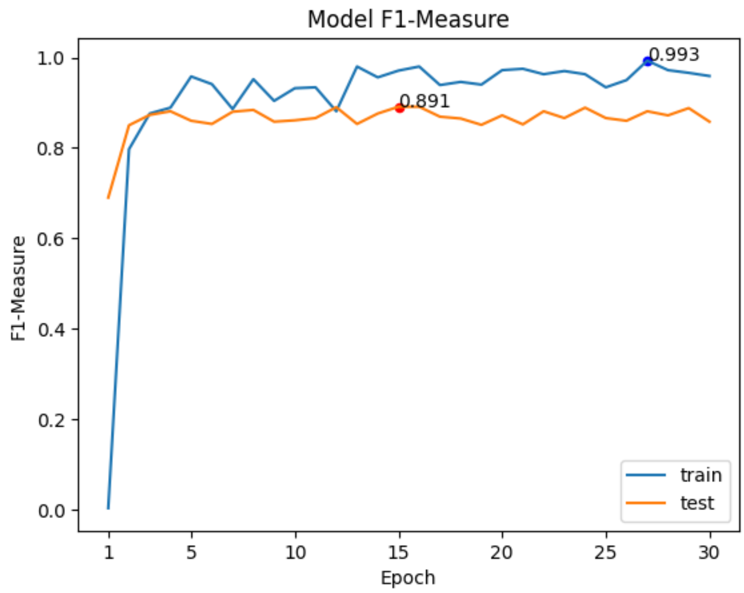

FIGURE 2. Base Line Model F1-Measure

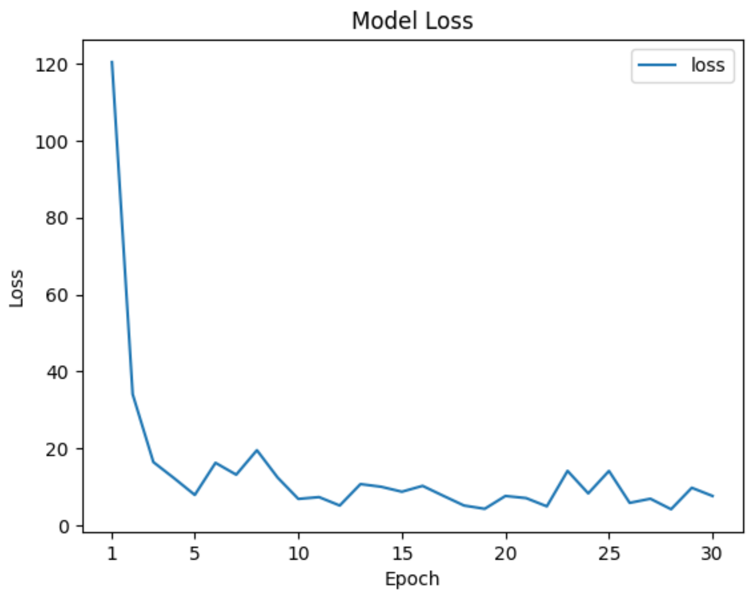

FIGURE 3. Base Line Model Loss

As we analyze in the methodology, the CRF layer of the baseline model fails to learn a sufficiently strong transition matrix when judging the constraint relationship of BIO. So, we choose to expand the learning rate of CRF. In our experiment, we chose to expand the learning rate of the CRF layer to 100 times and 1000 times that of the BERT layer. It can be seen from the Figure 4 that the F1-Measure of the model on the test set has a stable 1.9% increase, and it is also found that 1000 learning rate of 100 times is not much different from a learning rate of 100 times. By manually observing the prediction results of the test set, it can be found that the annotations that do not satisfy the lexical continuity are reduced, such as Begin-Drug should not be connected to Inside-symptoms.

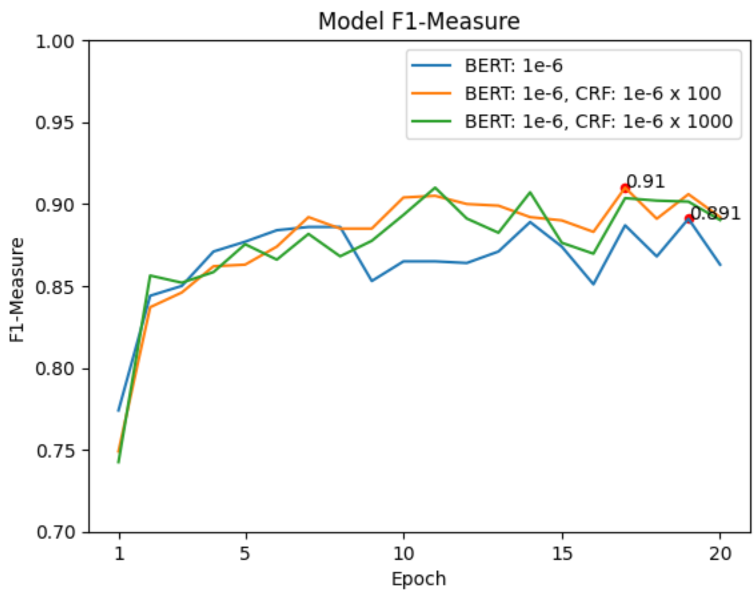

FIGURE 4. F1-Measure Comparison of Different Learning Rates

For missing and incorrect labels, this project uses a 10-fold label supplement solution. We divided the 6,000-piece training set into 10 pieces, and each time took 9 pieces of document, a total of 5,400 pieces of document for training. A total of 10 pieces were trained using the model after optimizing the learning rate above. Using these 10 models to predict the original training set, we can have 10 predicted training set results. When a label that appears in the original training set does not appear in any of the prediction results, we will remove this label. When all predictions are labeled with a valid label, but this label does not appear in the training set, we will add this label to the training set. Using this scheme, we can further optimize the training set with certain labeling problems when there is a lack of knowledge in the medical field.

We keep two updated datasets, one with only new annotations added and one with annotations added and removed. Using these two datasets to retrain the model, compared with the baseline model, we have achieved stable improvements in F1-Measure. Among them, the modified labeled dataset obtains the highest F1-Measure 92.7%, which is about 3.6% higher than the baseline model. Compared with the data set that only adds annotations, the data set that includes two operations of deletion and addition achieves better results. It means that if there is a problem with the label, blindly adding labels will not get better results, and deleting those data with errors can improve the recognition efficiency. Figure 5 shows a line graph of different training sets and different models on the F1-Measure evaluation.

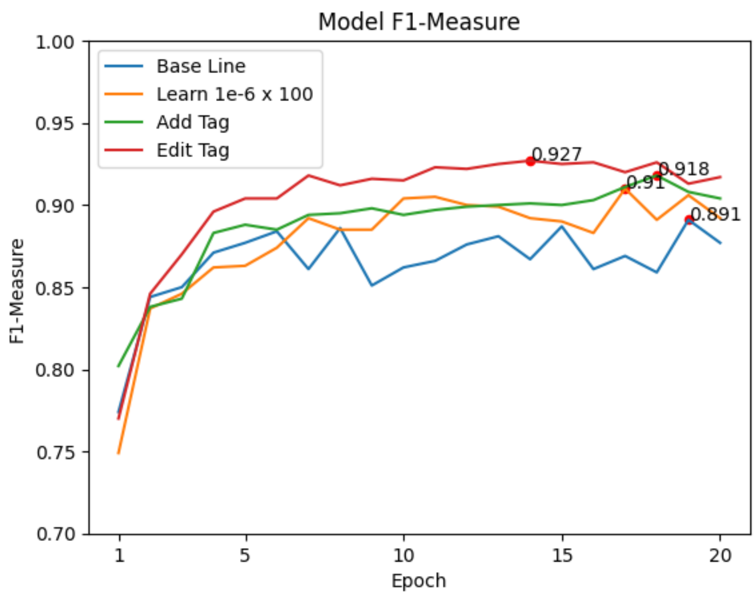

FIGURE 5. F1-Measure Comparison of Different Datasets

At the same time, by observing the loss curve of the model through Figure 6, it can be found that the modified data set can obtain faster and more stable loss indicators, and at the same time, it can be reduced to a lower level in the later stage of training, Achieved a loss value of 0.116. It means that the modified data set can obtain better fitting ability, which is in line with our expectation of reducing noise by processing the training set. In addition, by observing the accuracy rate and recall rate curves, it can be found that improving the learning rate of the CRF layer can slightly improve the accuracy rate, which is consistent with the role of the CRF layer in restricting label relationships in the model. The model trained using a 10-fold optimized training set can better improve the recall rate.

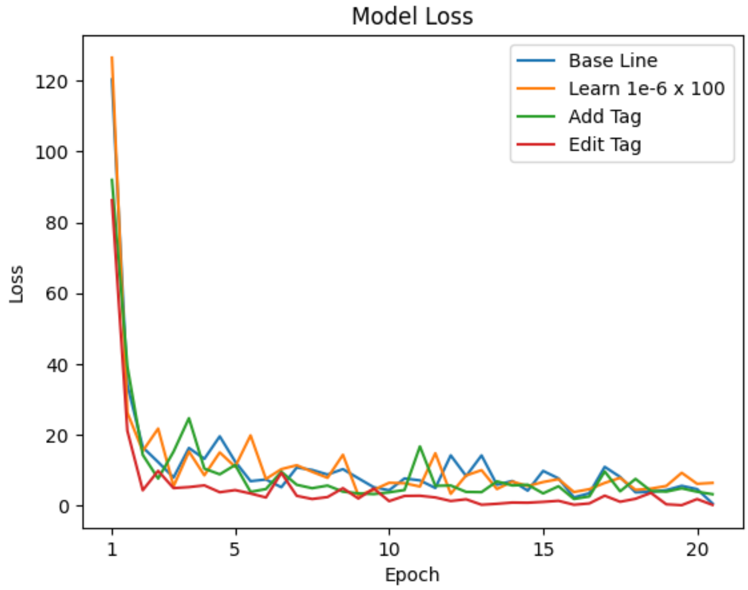

FIGURE 6. Loss Comparison of Different Datasets

By observing the accuracy rate and recall rate curves, it can be found that improving the learning rate of the CRF layer can slightly improve the accuracy rate, which is consistent with the role of the CRF layer in restricting label relationships in the model. The model trained using a 10-fold optimized training set can better improve the recall rate.

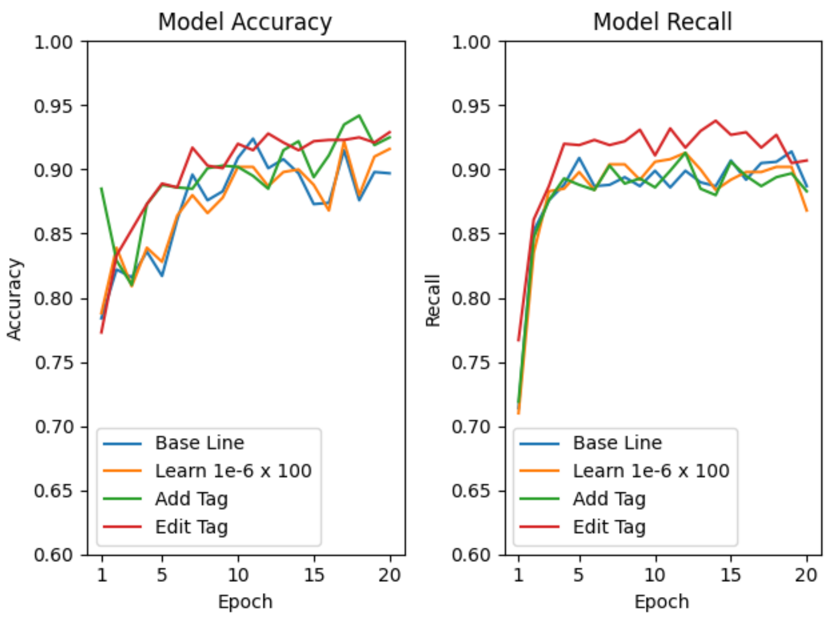

FIGURE 7. Model Accuracy and Recall

To summarize our experiments, in the baseline model, we first use 30 Epochs training to confirm that the model fitting ability does not improve after 20 Epochs, so we use 20 Epoch as a hyperparameter in subsequent experiments. The highest F1-Measure in the Add Tag training set appeared at the 18th Epochs, and the highest F1-Measure in the Edit Tag training set appeared at the 20 Epochs. This is due to the randomness included in the deep learning algorithm. It can be confirmed that the subsequent F1-Measure has been achieve stability.

TABLE 1. Best Performance Evaluation Summary

Model	F1-Measure	Accuracy	Recall	Loss
Baseline	0.891	0.866	0.917	0.58
Learning Rate x 100	0.910	0.922	0.913	4.44
Add Tag	0.918	0.942	0.913	2.56
Edit Tag	0.927	0.929	0.938	0.11

Accurate to the entity, the F1-Measure of the drug ingredient is the highest, reaching 98%. The efficacy of traditional Chinese medicine, diseases, symptoms and medicines reached 84%, 75%, 90% and 93% respectively. Among them, the recognition rate of diseases is slightly lower than that of other entities, and one of the reasons is that there are fewer disease labels in the original training set. Evaluation metrics for each entity are presented in Table 2.

Table 2. Summary of Performance Evaluations for Each Entity

Entity	F1-Measure	Accuracy	Recall
Drug Ingredients	0.98	0.98	0.98
Drug	0.93	0.94	0.93
Diseases	0.75	0.76	0.74
Symptoms	0.87	0.93	0.90
Traditional Chinese Medicine Efficacy	0.84	0.86	0.82

CONCLUSION

NER is an important basic tool in Knowledge graph and other application fields. It has played a crucial role in promoting the development of Natural language processing technology towards practicality. With the popularization of deep learning today, the effect of NER has been improved more universally than the previous dictionary-based method, but the improvement of NER still faces challenges in Chinese natural language, such as the problems caused by word segmentation and the improvement of the effect.

In order to improve the effectiveness of the Chinese dataset, this project uses the WanChuang dataset for model training and research. This is a NER dataset in the field of traditional Chinese medicine. In order to study more general methods, we limit the entity categories to 5: drugs, drugs, traditional Chinese medicine efficacy, symptoms, and diseases. This project proposes to use a pre trained model BERT+CRF as the baseline model, and two methods were proposed to achieve the research goal of improving NER extraction efficiency.

The first approach proposed in this study is to increase the learning rate of the CRF layer. We find that the learning rate of the pre-trained model BERT cannot learn enough constrained content for the CRF layer in a limited epoch. Since the CRF layer plays a role in limiting the rationality of the prediction results output by the BERT layer in the model structure we use, that is, the entity category should start with the BEGIN label, and the continuous entity categories should be the same. By increasing the learning rate of the CRF layer to 100 times that of the BERT layer, we obtained about 1.9% improvement in experiments. The highest F1-Measure reached 91%. It can be seen that the unreasonable label relationship has been significantly reduced by predicting the output of the result.

The second approach proposed in this study is to use a 10-fold training to optimize the training set. Through the analysis of the training set, we found that the training set contains more noise, including wrong labels and missing labels. Correcting these annotations manually requires not only strong expertise, but also unpredictable time and labor costs. In order to optimize the noise problem of the training set, we propose a 10-fold model optimization scheme. By dividing the training set into 10 parts and taking 9 parts each time to train a baseline model, we will obtain 10 completed models. Use 10 models to predict the training set and obtain 10 predicted training sets. When an entity annotation appears in the original training set but does not appear in any predicted training set, we delete the entity annotation; When an entity annotation does not appear in the original training set but appears in all predicted training sets, we will supplement the annotation of that entity. By using the modified training set to retrain the model, this experiment achieved the highest F1-Measure of 92.7%, which is about 3.6% higher than the baseline model, and an improvement of about 1.6% compared to the model that improved the learning rate mentioned above. At the same time, by observing the loss curve of the model, it can be found that the modified data set can obtain faster and more stable loss indicators, and at the same time, it can be reduced to a lower level in the later stage of training, Achieved a loss value of 0.116. It means that the modified data set can obtain better fitting ability, which is in line with our expectation of reducing noise by processing the training set.

In terms of future work related to this research, there are three parts that can be applied to expand the work of this project. The first part can effectively increase the exposure of deep learning models by integrating datasets from other Chinese medical fields. The second part is to optimize the model's ability in Chinese word segmentation by adding vocabulary in the medical field. The third part is due to the fact that some drugs in the medical field use their main components as their drug names. How to solve the problem of nested labeling is also the main focus of future work.

ACKNOWLEDGEMENT

I was lucky to have Dr. Saidah Saad as a mentor in my studies. To Dr. Saidah Saad, I would like to express my heartfelt thanks. She was an excellent and patient tutor who quickly answered and resolved my questions during my research and thesis writing.

REFERENCE

Bekoulis, Giannis, Johannes Deleu, Thomas Demeester, and Chris Develder. 2018. "Joint entity recognition and relation extraction as a multi-head selection problem."

Bin, Liu, Wu Zhiqiang, Wang Jianhong, Deng Lizong, Peng Yousong, and Jiang Taijiao. 2016. "Extracting Clinical entities and their assertions from Chinese Electronic Medical Records Based on Machine Learning." Atlantis Press.

Bui, Anita. 2019. "Traditional vietnamese medicine between chinese heritage and national tradition." *Chinese Medicine & Culture* 2 (1):21-5.

Chen, Xianglong, Chunping Ouyang, Yongbin Liu, and Yi Bu. 2020. "Improving the Named Entity Recognition of Chinese Electronic Medical Records by Combining Domain Dictionary and Rules." *International journal of environmental research and public health* 17 (8).

Cheng, Hu. 2014. "The Modernization of Japanese and Chinese Medicine (1914-1931)." *Chinese Studies in History* 47 (4):78-94.

Ju-Ah, Lee, Choi Tae-Young, Lee Myeong Soo, Ko Mimi, Kang Byoung-Kab, Liu Huan, Jiang Jun-Jie, and Li Yuan-Yuan. 2017. "Protocol for Systematic Review of Controlled Trials of Korean and Chinese Herbal Treatments for Stroke." *Journal of Acupuncture Research* 34 (4):169-71.

Krstev, Cvetana, Ivan Obradović, Miloš Utvić, and Duško Vitas. 2014. "A system for named entity recognition based on local grammars." *Journal of Logic & Computation* 24 (2):473-89.

Lei, Jianbo, Buzhou Tang, Xueqin Lu, Kaihua Gao, Min Jiang, and Hua Xu. 2014. "A comprehensive study of named entity recognition in Chinese clinical text." *Journal of the American Medical Informatics Association : JAMIA* 21 (5):808-14.

Luo, Ling, Zhihao Yang, Mingyu Cao, Lei Wang, Yin Zhang, and Hongfei Lin. 2020. "A neural network-based joint learning approach for biomedical entity and relation extraction from biomedical literature." *Journal of Biomedical Informatics* 103.

Roth, Adam, and Zhang Hongxia. 2016. "A Dialogue on Traditional Medicine: East Meets West." *China Media Research* 12 (4):85-92.

Sarkar, Chayna, Biswadeep Das, Vikram Singh Rawat, Julie Birdie Wahlang, Arvind Nongpiur, Iadarilang Tiewsoh, Nari M. Lyngdoh, Debasmita Das, Manjunath Bidarolli, and Hannah Theresa Sony. 2023. "Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development." *International Journal of Molecular Sciences* 24 (3):2026.

Tianchi. 2020. " Entity Recognition of Traditional Chinese Medicine Instructions."

Wang, Yu, Yining Sun, Zuchang Ma, Lisheng Gao, and Yang Xu. 2020. "Named Entity Recognition in Chinese Medical Literature Using Pretraining Models." *Scientific Programming* 2020:8812754.

Yu, T., J. Li, Q. Yu, Y. Tian, X. Shun, L. Xu, L. Zhu, and H. Gao. 2017. "Knowledge graph for TCM health preservation: Design, construction, and applications." *Artificial Intelligence in Medicine* 77:48-52-.

Zhang, Xuan, Lin Zhang, Jacky C. P. Chan, Xihong Wang, Chenchen Zhao, Ying Xu, Weifeng Xiong, Wai Chak Chung, Feng Liang, Xu Wang, Jiangxia Miao, and Zhaoxiang Bian. 2022. "Chinese herbal medicines in the treatment of ulcerative colitis: a review." *Chinese Medicine* 17 (1).