Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 11, No. 2, 2025

Research on the Construction of TCM Diagnosis Model based on Large Language Model

Mu Li, Yulin Xia, Wei Hu *, Ziyi Zhang and Chengchen Zhang

School of Informatics, Hunan University of Chinese Medicine, Changsha, Hunan, China
*** Corresponding author:** Wei Hu

Abstract: To solve many problems in TCM syndrome differentiation, including the lack of public data and quality differences, the problem of singleness and universality of models, and the lack of interpretability of models, a solution based on large language model ChatGLM3 combined with LoRA fine-tuning technology was proposed. The open source TCM-SD TCM syndrome differentiation data set was adopted, and after data filtering and integration optimization, 1027 TCM syndrome differentiation definition data sets, 41180 consultation training data and 5485 testing and verification data were obtained, so that the model could deeply learn the specialized knowledge of TCM syndromes and the actual consultation records of TCM syndrome differentiation. The experimental results show that the evaluation indexes of two different trainings using LoRA fine-tuning technology are significantly improved by about 20%.

Keywords: TCM Syndrome Differentiation; Large Language Model; ChatGLM3; LoRA Fine Tuning; Data Annotation.

1. Introduction

As a treasure of traditional Chinese culture, Chinese medicine possesses a set of perfect diagnostic theory system, which has made immeasurable contributions to the health and survival and development of human beings. The unique diagnostic and therapeutic methods of Chinese medicine have provided many valuable references for modern Chinese medicine, which has been warmly welcomed by people all over the world and has attracted extensive attention from medical researchers. As a precious resource of China's traditional medicine, TCM's method of evidence-based treatment plays a key role in clinical practice. However, the subjectivity and complexity of TCM evidence identification make the accuracy and consistency of TCM evidence identification and diagnosis face certain challenges. In 2022, the General Office of the State Council released the "14th Five-Year Plan for the Development of TCM"[1], which emphasises the integration of TCM with modern science and technology and aims to promote complementary and synergistic development between TCM and Western medicine, as well as to promote the development of TCM towards modernisation and industrialisation. In addition, the Plan encourages the development and application of information technology systems with Chinese medicine characteristics, such as the Intelligent Assisted Diagnosis and Treatment System for Chinese Medicine Diagnosis and Treatment [2] .

With the rapid development of AI technology, large-scale language models are gradually demonstrating their broad application prospects. These models have won the preference of many users because of their excellent performance in dialogue generation and semantic understanding, and their application scope is expanding to cover a wide range of industries such as education, finance, healthcare, law, entertainment and customer service. In 2023, "Qihuang Ask", a large model of Chinese medicine, introduced generative AI technology to the field of Chinese medicine for the first time, marking a groundbreaking progress in the research of large models of Chinese medicine in China. However, as the knowledge system of TCM is huge and covers many subfields, only relying on the basic macromodel is not enough to achieve in-depth interpretation of the knowledge of various subfields of TCM. Therefore, it is necessary to develop domain-specific large-scale models for TCM as an intermediary layer to better serve the TCM sub-domains [3]. Large-scale language modelling is a natural language processing technology based on deep learning, which has shown great potential for application as a knowledge aid for doctors in many aspects of medical clinics, research and education. Considering that Chinese medicine has a unique theoretical system and rich empirical knowledge, constructing a large language model for the field of Chinese medicine will provide an important opportunity to promote the intelligent and informative development of Chinese medicine.

2. Difficulties Faced by Chinese Medicine Practitioners in Identifying Evidence

2.1. Insufficient and Variable Quality of Publicly Available Data

Diagnosis in TCM involves a complex process of evidence identification and usually relies on the synthesis of multiple sources of information, such as patients' subjective descriptions, doctors' clinical observations, and the results of various examinations. However, the individualised, empirical and subjective nature of TCM diagnosis results in a relative lack of relevant public datasets. This lack of data directly affects the training and validation of intelligent discursive models, making it difficult for model performance to meet expected requirements. In addition, the emphasis on individualised treatment protocols in TCM means that the available public data may only cover a limited number of disease types or symptoms, lacking the necessary diversity and representativeness, which may affect the performance of the model when dealing with rare or special cases, limiting the scope of its practical application.

2.2. Single and Poorly Generalised TCM Diagnostic Models

Analysing the existing academic reports and practical application cases, the vast majority of intelligent evidence models are designed for a specific type of disease or a specific type of evidence [4]. This design approach leads to a significant limitation in the scope of application of the models, which makes it difficult to meet the demand for complex evidence identification processes in TCM clinical practice. On the one hand, this limitation stems from the constraints of the model design itself. Current intelligent discrimination models usually establish a fixed classification framework at the design stage, which means that diseases are preclassified into several fixed evidence types, and the models are trained based on these evidence types. Although this approach simplifies the problem processing flow to some extent, it introduces significant limitations at the same time. On the other hand, according to the viewpoint of Chinese medicine theory, the formation and progression of disease is a complex process that is influenced by a variety of factors intertwined with each other. However, current intelligent discursive models usually focus only on a specific few factors in this process, ignoring other factors that may have an impact on the diagnostic outcome. This simplification reduces the complexity of the model, but at the same time reduces the general applicability of the model. In a real clinical setting, doctors need to consider all factors comprehensively to conduct accurate evidence analysis, rather than relying solely on a few predefined and fixed evidence patterns [5].

2.3. Insufficient Interpretability of TCM Diagnostic Models

The process of discernment in Chinese medicine usually involves complex logical reasoning and rich practical experience, and current intelligent discernment models are significantly deficient in terms of interpretability. In addition, since most of these models are designed based on different schools of expert thinking, there are differences in interpretation between different models. Currently, the vast majority of intelligent TCM discernment studies rely on machine learning and deep learning models. However, due to the complexity and opacity of the internal workings of these models, it is difficult to explain their decision-making process, which often exhibits "black box" characteristics. This is difficult for non-specialists to understand. This makes it difficult for doctors to understand why the models make the judgements they do when outputting results, which reduces the acceptability of the models.

3. Construction of a Diagnostic Model for TCM

3.1. Base Models

Among many domestic open-source big models, Tsinghua University, in conjunction with Wisdom Spectrum AI, has attracted a lot of attention for the launch of its fully self-developed third-generation base big model, ChatGLM3, at the China Computer Congress (CNCC). Compared with the previous version, ChatGLM3 adopts diverse training data and reasonable training strategies, and shows excellent capabilities in multiple domains such as semantics, mathematics, reasoning, code, and knowledge. This comprehensive knowledge understanding and application capability provides a solid foundation for the TCM Big Model, making it more accurate and efficient in handling complex TCM knowledge quizzes. Compared with ChatGPT, ChatGLM3's full open source enables it to be deployed locally without the limitation of the number of APIs provided by enterprises. In addition, ChatGLM3 is specifically designed for dialogue tasks and is capable of generating smooth and natural dialogue responses. For the Chinese medicine big model, it is very important to be able to interact with users effectively and provide user-friendly consulting services. Considering the above factors, the private deployment of ChatGLM3 is chosen as the base model for this study to explore its application value in the field of TCM [6].

3.2. Datasets

The data source of this article is AliCloud Tianchi open source TCM-SD TCM domain identification data (https://tianchi.aliyun.com/dataset/139034), TCM-SD is the first open, real-scene collected TCM identification dataset, which aims to solve the TCM identification problem by using natural language processing technology, and explore the latent scientific basis behind the TCM identification theory, based on the existing dataset, TCM-SD performs a secondary optimisation of the data through the two phases of commanded data filtering-commanded data integration, and the optimisation of the data involves The optimisation of data involves multiple steps, including but not limited to data collection, data preprocessing, and data annotation. The general steps for constructing such a dataset are shown in Figure 1.

Fig 1. Flowchart of dataset construction

The dataset is divided into training, validation and test sets, and the division ratio used is 80% of the data for training, 10% for validation and 10% for testing. Finally, the dataset is formatted and divided into training dataset of 41180 entries and test and validation sets of 5485 entries, where some of the training data examples are shown in Table 1.

In order to improve the understanding and application of large-scale models in the field of TCM in order to provide more accurate results, this study collects a total of 1,027 TCM evidence definitions datasets, of which some of the TCM evidence definitions are shown in Table 2. Through in-depth learning of the TCM domain-specific evidence expertise, the model is better able to adapt to specific tasks and scenarios, thus demonstrating higher efficiency and accuracy in practical use. The pre-acquired knowledge of TCM enables the model to better meet user needs and provide more personalised and valuable information.

Table 1. Some examples of training data

Symptomatic	Syndrome Differentiation	TCM Syndrome
The patient had black stools seven days ago due to poor diet, which was carried out once a day with small volume and no mucus, accompanied by paroxysmal dull pain in the abdomen, no radiating pain in the lower back and lower limbs, acid reflux, belching and burning sensation behind the sternum, no nausea and vomiting, no fever, fatigue, no palpitation and chest tightness, poor food intake, restless sleep and irregular urination.	Clear mind, reasonable spirit, moderate body, clear language, red lips; Normal skin, no macular rash. Neck symmetry, no exposed veins, no scrofula of gall, chest symmetry, normal pulsing in the hollow, flat abdomen, no mass, red nail color, no edema of the lower extremities, red tongue, thin yellow fur, string pulse.	Syndrome of Qi deficiency and blood overflow
One month ago, the patient felt anal knife pain and discomfort due to constipation, and the anal mass was propped out after the stool, which could not be recovered, and the stool was bloody. The symptoms did not see significant relief after resting on her own, and the symptoms became significantly worse 3 days ago, with severe anal pain, difficulty in walking, and excessive blood volume in the stool, so today the outpatient department proposed "mixed hemorrhoids, anal fuses, and anal papilloma.	Clear mind, reasonable spirit, moderate body, clear language, red lips; Normal skin, no macular rash. Neck symmetry, no exposed veins, no scrofula of gall, chest symmetry, normal pulsing in the hollow, flat abdomen, no mass, red nail color, no edema in the lower extremities, pale red tongue, white fur, pulse calm.	Dampness-heat syndrome
The patient had edema of both lower limbs 33 years ago, and went to the hospital for treatment. Several urine routine examinations showed proteinuria positive, and was diagnosed as chronic nephritis. She had taken Chinese medicine for treatment, but had not been given regular treatment, and edema of both lower limbs occurred repeatedly. Palindrome catheter was inserted into the tunnel. The internal fistula was mature. The right femoral vein was removed and the central vein was placed in hospital.	Clear mind, reasonable spirit, moderate body, clear language, pale lips; The skin is pale without macular eruption. Normal skull size and shape, no eyes depression, no white eyes, normal ear rings, no ear fistula or sores; Neck symmetry, no exposed veins, no scrofula of gall, chest symmetry, normal pulsing in the hollow, flat abdomen, no mass, red nail color, no edema of the lower extremities, pale red tongue, white fur, deep pulse.	Spleen-kidney deficiency syndrome

Table 2. Examples of some Chinese medicine evidence definition data

Name	Definition	Typical characteristic expression	Common disease
Syndrome of wind-cold attacking lung	It refers to the common syndrome of wind-cold invasion, lung qi loss, aversion to cold, no sweating, cough, chest tightness, white phlegm, white fur, pulse floating tight and so on.	Cough, phlegm clear thin, stuffy nose, runny nose, throat itching, both cold, light fever, thin white tongue coating, pulse floating tight.	Cough, infantile bronchitis, acute trachea-bronchitis
Heart-kidney Yang deficiency syndrome	Refers to the heart and kidney Yang deficiency, loss of warm, cold limbs, palpitation, poor urination, swollen limbs, cold waist and knees, pale purple tongue, white fur, weak pulse and other common symptoms.	Cold limbs, palpitation, palpitation, chest tightness, asthma, swelling of the limbs, unfavorable urination, fatigue, cold waist and knees, blue lips, pale tongue, smooth white fur, weak pulse.	Edema, palpitation, palpitation, chest numbness
Syndrome of wind-heat blocking lung	External invasion of wind heat, lung qi stagnation, fever and wind aversion, cough, rough breath and panting, chest tightness and chest pain, nose flaring, no sweat, red tongue, floating pulse and so on are the common symptoms.	Fever, ill wind, cough, rough breath, chest tightness, chest pain, flaring nose, no sweat, red tongue, floating pulse.	Pneumonia cough, neonatal dyspnea

3.3. Fine-tuning of the Model

LoRA (Low-Rank Adaptation) [7] is an efficient model fine-tuning technique that aims to reduce the number of parameters required for fine-tuning by inserting low-rank matrices into a pre-trained model, thus improving training efficiency and avoiding overfitting. The core idea of LoRA is to keep most of the parameters of the pre-trained model unchanged, and simulate the parameter changes by adding low-rank matrices. amount of variation, thus achieving adaptation to a specific task. In this study, the fine-tuning using LoRA is divided into two phases, firstly, the first phase of fine-tuning for the above TCM evidence data set to learn the knowledge of TCM evidence definition, and secondly, the second phase of fine-tuning through the above training set of TCM diagnosis patients' case consultation, in which the parameters of the fine-tuning for LoRA are shown in Table 3.

Table 3. Parameters for LoRA fine-tuning

Fine-tuned parameters	Parameter value or type
per_device_train_batch_size	64
gradient_accumulation_steps	1
per_device_eval_batch_size	8
num_train_epochs	5
learning_rate	10e-4
lr_scheduler_type	cosine
warmup_steps	20
weight_decay	0.01
lora_rank	16
lora_alpha	8
lora_dropout	0.05

The number of iterations (Step) refers to the process of using a batch of data for one parameter update during the training process of the model. In deep learning, in order to efficiently use the computational resources and improve the training speed, the entire dataset is usually not input into the model at once for training, but rather, the dataset is split into multiple small batches. For each Batch processed, the model performs a forward propagation, calculates the loss, backpropagation, and update the parameters. Loss Function (Loss Function) is a computational tool used to quantify the difference between the predicted and actual values of the model, with smaller values indicating higher robustness of the model. The function is mainly applied in the training phase, when each batch of data is fed into the model and after forward propagation produces predictions, the loss function calculates the error between these predictions and the true labels, known as the loss value. Based on this loss value, the model adjusts the internal parameters through the back-propagation mechanism to reduce the prediction error and drive the prediction results closer to the real situation, thus achieving the learning objective. From Figures 2 and 3, it can be observed that the loss value gradually decreases with the increase in the number of training iterations and stabilises after reaching 90 iterations, indicating that the model has basically completed the learning process.

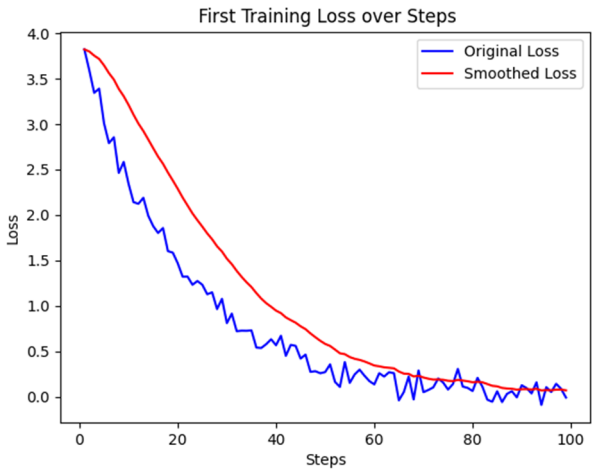

Fig 2. Graph of step loss for the first training **session**

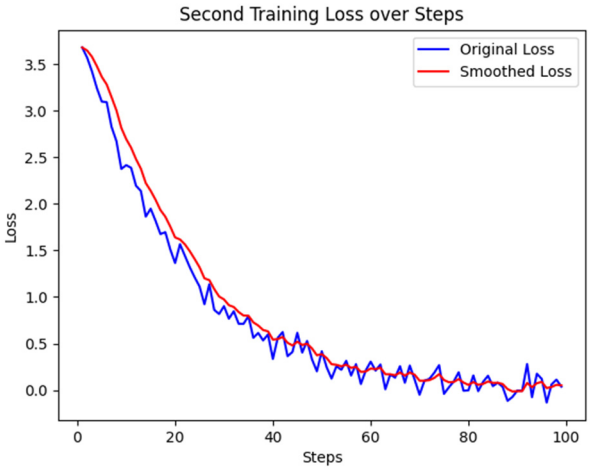

Fig 3. Graph of step loss for the second training session

3.4. Fine-tuning of the Model

Automatic evaluation of language models is a common and popular evaluation method, mainly because it saves a lot of time and cost by not requiring human involvement, and it usually uses standard metrics or indicators to evaluate the performance of the model, such as BLEU (Bilingual Evaluation Understudy) [8], ROUGE (Recall- Oriented Understudy for Gisting Evaluation scores) [9] and so on. Among them, BLEU is an evaluation metric for machine translation tasks, the core idea is to compare the degree of overlap between the N-grams in the generated text and the reference text, and the higher the degree of overlap, the better the quality of the translated text is. ROUGE metrics are similar to the BLEU metrics, except that ROUGE is based on the recall rate, while BLEU is more concerned with the accuracy rate [10] . ROUGE-N emphasises on the completeness of the generated text, while ROUGE-L tends to take into account the accuracy rate. while ROUGE-L tends to consider the completeness and order of sentences. In this study, BLEU metrics (generally N selects 4) and ROUGE metrics are used to automatically evaluate the validation dataset, with the value of 0-1 points, and the higher the value means the better the quality of the output. The specific scores of each indicator are shown in Table 3 below. From the results, the generation effect of fine-tuning the large model is significantly better than that of not fine-tuning, and the generation effect is significantly improved by the second fine-tuning training, and the values of the overall indicators are all improved by about 0.2.

Table 4. Score table for model evaluation indicators

Model	BLEU-4	ROUGE-1	ROUGE-2	ROUGE-L
ChatGLM3	0.6261	0.6656	0.4461	0.6078
ChatGLM3+LoRA (first train)	0.6434	0.7316	0.5493	0.6174
ChatGLM3+LoRA (second train)	**0.8467**	**0.9319**	**0.6639**	**0.8172**

4. Summary

This paper first expounds the importance of TCM and its significant contribution to human health. At the same time, it points out the three major challenges faced by TCM diagnosis and differentiation: insufficient and inconsistent quality of public data, poor uniformity and universality of diagnostic models, and insufficient interpretability of models. To solve these problems, the large model ChatGLM3 is used as the infrastructure in this study, and efficient LoRA fine-tuning technology is introduced to optimize the model performance by using two training sessions. The results showed that BLEU-4 score increased from 0.6261 to 0.8467, ROUGE-1 score increased from 0.6656 to 0.9319, ROUGE-2 and ROUGE-L score also increased from 0.4461 and 0.6078 to 0.6639 and 0.8172, respectively. The model fine-tuned by LoRA has a significant improvement in various indexes. It is

proved that the model can significantly improve the semantic understanding and the generation of dialectical diagnosis. Through this study, it not only provides a more accurate and efficient intelligent auxiliary tool for TCM syndrome differentiation and diagnosis, but also takes an important step for the intelligent and information process of TCM.

Acknowledgments

This research was supported in part by Hunan University of Chinese Medicine Undergraduate Research and Innovation Fund Project (No: 2023BKS081); 2024 the National College Student' Innovation and Entrepreneurship Training Program Project (No: S202410541140).

References

[1] General Office of the State Council of the People's Republic of China. The 14th Five Year Plan for the development of traditional Chinese medicine [EB/OL]. (2022-03-03)[2024-02-03]. https://www.gov. cn/gongbao/ content/ 2022/ content_ 568 6029.htm.

[2] Wang Ye. Research on Intelligent Chinese Medicine Discriminatory Method Based on Deep Learning [D]. Henan University of Science and Technology,2022.000120.

[3] SONG Yijie, MA Suya, DAI Yasheng, et al. Key issues and technical challenges of artificial intelligence-assisted Chinese medicine discernment[J]. China Engineering Science,2024,26 (02): 234-244.

[4] YANG Lele, WANG Zhe, YAO Keyu, LIU Lihong, ZHU Yan. Prospective thoughts on the application of big language modelling in the field of traditional Chinese medicine[J/OL]. Chinese Journal of Traditional Chinese Medicine, 1-20.

[5] Wang R, Pan C, Chen J, et al. Construction of a knowledge framework system for intelligent diagnosis in Chinese medicine [J]. Journal of Traditional Chinese Medicine, 2024, 65(4): 341-346.

[6] LIU Yuehan, HUO Haobin, JIN Changuo. Practice and exploration of building enterprise-level private big language model assistant based on ChatGLM3 and RPA technology[J]. Architectural Design Management,2023,40(12):33-40.

[7] Hu E J , Shen Y , Wallis P ,et al. LoRA: Low-Rank Adaptation of Large Language Models[J]. 2021.DOI: 10. 48550/ arXiv. 2106.09685.pdf.

[8] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation [C]// Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002:311-318.

[9] LIN C Y. Rouge: A package for automatic evaluation of summaries[C]//Text summarisation branches out. 2004:74-81.

[10] Zhang Jundong, Yang Songhuah,Liu Jiangfeng,et al.AIGC empowers the revitalisation of ancient Chinese medicine:the construction of the Huang-Di grand model[J].Library Forum, 2024, 44(10):103-112.